FURTHER STUDIES ON THE MEMORY FACTOR

BY
ANNE ANASTASI, Ph.D.
Instructor in Psychology
BARNARD COLLEGE

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor
Columbia University

No. 142

NEW YORK
August, 1932

ARCHIVES OF PSYCHOLOGY

COLUMBIA UNIVERSITY, NEW YORK CITY

The Subscription price is six dollars per volume of about 500 pages. Volume I comprises Nos. 2-10; Volume II, Nos. 11-18; Volume III, Nos. 19-25; Volume IV, Nos. 26-32; Volume V, Nos. 33-39; Volume VI, Nos. 40-46; Volume VII, Nos. 47-52; Volume VIII, Nos. 53-58; Volume IX, Nos. 59-63; Volume X, Nos. 64-68; Volume XI, Nos. 69-73; Volume XII, Nos. 74-78; Volume XIII, Nos. 79-85; Volume XIV, [illegible]

FURTHER STUDIES ON THE MEMORY FACTOR

BY
ANNE ANASTASI, Ph.D.
Instructor in Psychology
BARNARD COLLEGE

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor
Columbia University

No. 142

NEW YORK
August, 1932

2788

TABLE OF CONTENTS

Chapter — *Page*

I. Introduction 5
   1. Setting of the Problem 5
   2. Plan of the Present Study 7

Part I

Verification of the Author's Previous Study

II. Procedure 9
   1. The Subjects 9
   2. The Tests 11

III. Results 13
   1. Evaluation of the Tests 13
   2. Intercorrelations of the Variables 14
   3. The Tetrad Analysis 18
   4. Correlations with Central Factor and Regression Equation 21
   5. Analysis of Variance 23
   6. Conclusions 24

Part II

The Question Examined over a Wider Range of Memory Tests

IV. Procedure 26
   1. The Subjects 26
   2. The Tests 26
   3. Preliminary Experimentation 29
   4. Testing Procedure 32
   5. Scoring 36

V. Results 38
   1. Evaluation of the Tests 38
   2. Intercorrelations of the Variables 39
   3. Conclusions 45

Part III

Recall and Recognition

VI. An Analysis of Logical Recall and Recognition 47
   1. Construction of the Tests 47
   2. Procedure 50
   3. Results and Conclusions 52

VII. Summary and Evaluation 55

VIII. Bibliography 59

Further Studies On the Memory Factor

Chapter I

INTRODUCTION*

I. *Setting of the Problem*

The present series of studies center about the problem of whether memory may be considered a unitary and differentiable mental trait. The investigations reported in this monograph were undertaken as a continuation and an elaboration of a preliminary study on the memory factor published in 1930(2). The procedure and results of the present study should be considered throughout in the light of the earlier study. The reader is referred to the 1930 study for an analysis and critique of the literature bearing on the general problem. It was pointed out in that study that it is very difficult to draw any definite conclusions from the literature on the memory factor because of various limitations in the techniques employed by other investigators. The results obtained in the 1930 study yielded several lines of evidence suggesting the presence of a single common factor through tests of immediate rote memory for visually presented material. Furthermore, this memory factor varied quite independently of performance in three other tests used, which were taken as measures of verbal,[1] numerical,[2] and spatial[3] ability, respectively. The data in that study were secured exclusively on a group of male college students, nearly all of whom were in the junior year. These results were therefore limited, first, from the standpoint of the tests used, and secondly, from the standpoint of the type of subjects on whom they were established. Insofar as the tests and the subjects constituted random samplings of

---

* The writer wishes to express her appreciation to Prof. H. E. Garrett of Columbia University for helpful suggestions and criticisms, as well as for having allotted a portion of a grant from the Columbia University Council for Research in the Social Sciences to cover the larger portion of the statistical and clerical expenses of this study.

[1] Vocabulary Test used by Schneck (26).
[2] Arithmetic Reasoning Test used by Schneck (26).
[3] The Minnesota Paper Form Board Test (3).

certain populations, in the statistical sense, we can generalize our conclusions to cover the entire populations. Since our tests represented a fairly wide and random selection of immediate rote memory tests for visual material, we may conclude that our results indicated the presence of a common factor through *such tests* in general, but not necessarily through any other varieties of memory tests not sampled in our study. Likewise, our subjects were fairly representative of male college students of a certain specified social and racial status (cf. 2, p. 26-27). We cannot assume, however, that the same trait relationships would hold for a group differing in sex, age, race, social or educational status from our experimental group. It is quite justifiable—in fact, fundamental in most statistical work—to generalize from the experimental sampling to the population from which it is drawn, but that population should itself be clearly defined and delimited, and the conclusions should not be extended indiscriminately to other populations, without experimental verification.

A further limitation inherent in any single statistical study of trait organization is that the results throw little or no light upon the *nature* of the common factor discovered. This is especially true when the range of functions tested is fairly narrow and the tests used are similar in more than one respect. Although the presence of a common factor may be established mathematically, it is usually difficult to determine just what it is that is common to all the tests concerned. For example, the common factor might result from special skills and techniques acquired during the course of general training and education. All the tests in which one of these special techniques or "tricks of the trade" could be applied would show a common factor. One such technique is the well-known device of forcing meaningful associations in rote memory tests. Subjects probably differ considerably in the readiness and appropriateness of such associations, a fact which *might* account to a large extent for the differences in performance on such tests. Common devices such as this could easily produce a common factor by themselves. A second explanation that suggests itself is in terms of community of *material*, or content, of the tests used. Some subjects may display greater proficiency than others in dealing with a given type of material, irrespective of what is to be done with that material. Thus, an individual

especially adept in mechanics might excel in mechanical construction, perceiving mechanical relations, and recalling mechanical facts and concepts. Such special proficiencies in dealing with a certain type of material might in turn have resulted from any number of factors, such as innate capacity, training, interest. A third explanation which may be offered to account for a mathematically established common factor is that the tests involved are all based on a common mental *process*, a unitary trait, which is manifested to a greater or lesser degree whatever the material may be. Other explanations could no doubt be suggested. All too often the third explanation is assumed to be the correct one, although there may be no direct evidence for it in the data. By repeating an investigation on different types of subjects, and using different combinations of tests covering a wide range of material, it should be possible to arrive at a somewhat clearer understanding of the nature of the common bond.

II. *Plan of the Present Study*

The present study is divided into three parts, each part representing a separate investigation. In Part I, the results of the 1930 study were checked on a different population. The subjects were college women, one year younger on the average than the college men used in the earlier study, and the majority were sophomores rather than juniors. In addition, it was possible in the present study to analyze more fully the relation between memory and the verbal and numerical factors, since *two* verbal and *two* numerical tests were used, rather than one of each as in the earlier study. Tetrads could therefore be computed with two memory tests combined with two tests of one of the other abilities, and the independence of the group factors involved could be demonstrated. In the 1930 study, it was not possible to demonstrate the independence of the memory factor by means of the tetrad criterion, but only by the size of the correlations themselves. Finally, the data in Part I of our study offer a means of checking on a different population some of the results on the verbal and numerical factors, secured by Schneck (26), since some of the original tests used by Schneck were repeated on our subjects. Part I may, therefore, be characterized in general as a checking over of results formerly obtained, by varying the type of

subjects. No new tests were used, and the procedure was purposely kept identical to that previously followed.

In Part II, the *tests* themselves were varied. Our main purpose now was to find the extent of the memory factor and to throw some light, if possible, on its nature. Can the memory factor previously found be regarded as sheer retentivity, so that it will be manifested in any test involving retention, or is some of the overlapping found due to similarity of material, special techniques, etc.? The tests used in this part of the study differed as much as possible from each other and from the earlier tests, the only common feature through all of them being the fact that the subjects were required to *retain* certain impressions, in each test.

Part III is a special analysis of logical recall and recognition. In any analysis of the memory factor, the relation of these two processes should be considered. They have frequently been discussed as two separate processes, and statements have been made as to their relationship, their relative difficulty and the differences in their susceptibility to various factors.[4] In this study, our aim was to make the tests of recall and recognition as comparable as possible, so as to eliminate the effect of any extraneous factors.

The implications for the memory factor of all the three studies are brought together and analyzed in Chapter VII. Hypotheses regarding the extent and nature of the memory factor are offered, in the light of the results secured through the various attacks on the problem in Parts I, II, and II.

---

[4] Cf., for example, Strong (30, 31) and Hollingworth (11).

PART I*

Chapter II

PROCEDURE

I. *The Subjects*

The subjects in this study were 186 women students at Brooklyn College, enrolled in five sections in introductory psychology, the individual sections containing from 29 to 42 students. All subjects whose records on any one test had to be omitted because of absence, misunderstanding directions, or any other irregularity of procedure, were excluded entirely from the final experimental group. This left a remainder of 140 subjects on whom are based all the data reported in this study. The group was very homogeneous: nearly all the subjects were sophomores; the large majority were American-born of Russian Jewish parentage, and came from middle class homes. Age variability was very slight, but in order to reduce further its effect on individual differences in test performance, it was held constant in all the correlations by the use of the partial correlation technique. Data on the homogeneity of our group were secured by having each subject fill out a questionnaire, which is reproduced below.

DATA SHEET—A—

Name .... Class................

Age: Years..........Months..........Place of Birth................

Mother's Birthplace..............Father's Birthplace...............

Is any language other than English spoken at home?..................

Father's Occupation ....

Father's Education:

Elementary School.......................................

High School..

College ....

Are you Jewish?...

Do you have any part-time employment while at college?.............

If so, how many hours per week does it occupy?......................

The results of this questionnaire are summarized in Table I. All the figures in that table are expressed as *percentages*,

* The writer is indebted to Dr. S. E. Asch of Brooklyn College for his cooperation in obtaining the subjects and for administering the tests in this part of the study.

Table I

ANALYSIS OF SUBJECTS

Class		*Age*	*Birthplace*		*Father's Birthplace*	*Mother's Birthplace*
Low Junior	6	Mean = 18.61	United States	96	19	33
Sophomore	92	S.D. = 1.34	Russia and Poland	3	49	39
Upper Freshman	2	Range: 16.33 to 28.25	Austria	1	18	20
		Sk/σ_{Sk} = —0.85	Germany	0	3	1
		Ku = .2106 ± .0235	Italy	0	6	5
			Norway and Sweden	0	3	2
			British Isles	0	2	0

Jewish		*Other Language Spoken at Home*		*Father's Education*		*Father's Occupation*		*Hours of Outside Employment per Week*	
Yes	77	Yes	60	None in U. S.	20	None Specified*	10	None	72
No	23	No	40	Elementary School	51	Skilled Labor	30	1– 4	5
				High School	25	Merchants, Shopkeepers, and Clothiers	38	5– 9	16
				College or Professional School	4	Clerical and Selling	15	10–14	3
						Professional and Semi-Professional	7	15–19	2
								20 and Over	2

*Includes: deceased, retired, and unemployed.

except in the case of age, in which the mean, standard deviation, and range of the entire distribution, as well as measures of skewness and kurtosis, are reported. The measure of skewness is that given by Kelly (13—p. 77) and reported, with a correction in the formula for its standard error, by Dunlap and Kurtz (6—p. 112). A positive skewness with this formula denotes a piling up of measures at the upper end of the distribution, a negative skewness, a piling at the lower end. The measure of kurtosis is also taken from Kelly (13—p. 77). A value of .26315 obtained by this formula represents the kurtosis of the normal curve. It will be seen that the age distribution of the subjects in this study conforms closely to the normal curve.

II. *The Tests*

The tests used include four tests of memory, two tests of verbal ability, and two of numerical ability. The four memory tests were selected from the 1930 study as being the four *single* tests[1] which most nearly satisfied the tetrad criterion when taken together. They were the word-word paired associates, the picture-number paired associates, the nonsense syllable recognition, and the retained members for words tests (for full description, see (2)). The verbal and numerical tests were taken from Schneck's study (26). They included a vocabulary test and an analogies test (both given by the multiple choice method), an arithmetic reasoning test, and a number series completion test. These tests were selected because they gave the highest correlations with the verbal and numerical factors, respectively. Exactly the same procedure was followed in giving the tests as in the earlier studies. Revisions and modifications were purposely avoided so that the present investigation might offer a direct check on the former results.

The tests were all given by one examiner, the regular instructor in the classes used. The procedure was kept standard by typewritten directions for the experimenter, standard mimeographed blanks for the subjects in all the tests, and uniform timing, controlled by stop watch or metronome. The testing extended over a period of four months during the first semester of the academic year 1931-32; it required a total of three hours and forty minutes, and was scattered over six

[1] I.e., not pooled tests.

TABLE II

TESTING SCHEDULE

Period	*Tests, in the Order Given*	*Number of Items*	*Approximate Duration of Entire Test, in Minutes**
I.	Vocabulary	130	40
II.	Analogies	40	35
III.	Number Series Completion	41	35
IV.	Arithmetic Reasoning	40	50
V.	Word-Word	40	15
	Picture-Number	40	15
VI.	Retained Members	40	10
	Syllable Recognition	80	20

*This does not mean time limit given to the subjects, but total time required for the test.

class periods. The testing schedule, together with the number of items and duration of each test, are reproduced in Table II.

All the tests were scored in terms of total number of items correct. No partial credits were used. The highest possible score in each test is equal to the total number of items in that test, as given in Table II. The score in the nonsense syllable recognition test was taken as total number of items minus twice the number of wrong items; no omissions were allowed in this test.

Three scores were computed for each test, viz., the total score, the score in the "odd" items, and the score in the "even" items. Reliability coefficients were computed by correlating odds and evens scores and applying the Spearman-Brown formula to find the reliability of the entire test.

Chapter III

RESULTS

I. *Evaluation of the Tests*

All of the data reported in the present chapter are based upon the 140 subjects whose records were complete. Before considering the intercorrelations of the variables, the tests themselves will be examined with regard to variability, reliability, and normality of distributions. In evaluating a test, the variability of the distribution of test scores should be considered. When the variability of the test is narrow, differentiation is often poor, unless the test units are small. Moreover, poor differentiation lowers the reliability coefficient of a test, as well as its correlation with other tests. Tests yielding a high variability are, therefore, desirable in studies in differential psychology. The normality of the distribution of test scores may be taken as an index of how well the test is adapted to the particular group of subjects tested. If, for example, a test is too easy for the subjects, there will be a piling up of scores at the upper end, yielding a positively skewed distribution; if the test is too difficult, the distribution will be negatively skewed, with the majority scoring at the lower end.

In Table III will be found the reliability coefficients, means, and standard deviations, as well as measures of skewness and kurtosis, for each variable. The tests have been numbered in the order in which they were given, and will be referred to henceforth by these numbers. The reliabilities of the tests are fairly satisfactory. They range from .64 to .93, and agree quite closely with those previously obtained with the same tests. The one exception is the reliability coefficient of the analogies test. This is .702 as compared with the reliability of .879 reported by Schneck (26). But in view of the fact that the variability of this test is very low—lower than that of any of the other tests, in fact—a very high reliability coefficient would be rather surprising. We should hardly expect, for example, that it have as high a reliability as the arithmetic reasoning test (.891), which is a longer test and has a much higher variability.

TABLE III

EVALUATION OF THE TESTS*

Test	*Reliability Coefficient*	*Mean*	*Standard Deviation*	Sk/σ_{Sk}	*Ku.*
1. Vocabulary	.9193	61.5643	16.6494	—0.68	.2310
2. Analogies	.7022	24.8500	4.0934	—2.63	.2288
3. Number Series Completion	.9317	30.4786	7.1087	+2.85	.2670
4. Arithmetic Reasoning	.8910	22.5071	7.2719	—2.01	.2717
5. Word-Word	.8634	17.3786	6.7182	—0.23	.2477
6. Picture-Number	.8331	14.7071	6.1732	—3.23	.2649
7. Retained Members	.6608	19.6786	4.3543	—0.63	.2777
8. Syllable Recognition	.6369	46.7000	10.2342	—0.26	.2761

*The means, standard deviations, and correlations of halves were computed by the Columbia University Statistical Bureau.

In all the tests except number series completion, there is a tendency towards negative skewness, suggesting that the tests proved to be somewhat too difficult for our subjects. In no case, however, is the skewness significant except in the picture-number test, where Sk is 3.23 times as large as σ_{Sk}. Even in this case, the skewness probably has little effect on the results. The data obtained with this test do not show any appreciable difference from those obtained with the other tests, but seem to behave in very much the same way. The measure of kurtosis used, (Q/D), yields a value of .26315 for a normal curve. The standard error of this measure with 140 cases is $\frac{.27779}{\sqrt{140}}$, or .0235. It is clear from Table III that none of our distributions shows a significant deviation from normal kurtosis.

II. *Intercorrelations of the Variables*

In Table IV will be found all the intercorrelations among the eight tests used. Raw coefficients are shown above the diagonal extending from the upper left to the lower right hand corners; corrected coefficients are below the diagonal. Table V gives the partial correlation coefficients, with age variability held constant, corresponding to the intercorrelations in Table IV.

The correlations of each test with age are as follows:

1. Vocabulary	.1460
2. Analogies	.1478
3. Number series completion	—.1442
4. Arithmetic reasoning	—.1606
5. Word-word	—.0199
6. Picture-number	.0434
7. Retained members	—.1193
8. Syllable recognition	—.1126

TABLE IV

INTERCORRELATIONS OF THE VARIABLES*

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1. Vocabulary		.6588	.0918	.2040	.0246	.0716	—.0338	.1301
2. Analogies	.8200		.2202	.2932	.1841	.0512	.0049	.2787
3. Number Series Completion	.0992	.2723		.5902	.1021	.1569	.2255	.0635
4. Arithmetic Reasoning	.2254	.3707	.6480		.0929	.1414	.2549	.2446
5. Word-Word	.0276	.2364	.1139	.1059		.5075	.6058	.3273
6. Picture-Number	.0818	.0669	.1782	.1641	.5984		.4576	.2337
7. Retained Members	—.0434	.0072	.2875	.3322	.8021	.6167		.2714
8. Syllable Recognition	.1700	.4168	.0825	.3247	.4414	.3208	.4183	

*Raw correlation coefficients above the diagonal, corrected below.
The raw correlation coefficients were computed by the Columbia University Statistical Bureau.

TABLE V

PARTIAL CORRELATION COEFFICIENTS, WITH AGE VARIABILITY HELD CONSTANT*

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1. Vocabulary		.6513	.1153	.2329	.0278	.0661	—.0167	.1490
2. Analogies	.8153		.2468	.3246	.1891	.0453	.0229	.3005
3. Number Series Completion	.1248	.3069		.5805	.1003	.1651	.2120	.0481
4. Arithmetic Reasoning	.2580	.4131	.6390		.0909	.1505	.2405	.2310
5. Word-Word	.0313	.2441	.1120	.1038		.5090	.6078	.3273
6. Picture-Number	.0756	.0595	.1876	.1749	.6002		.4666	.2404
7. Retained Members	—.0215	.0340	.2715	.3151	.8080	.6313		.2615
8. Syllable Recognition	.1955	.4528	.0627	.3079	.4425	.3308	.4056	

*Raw correlation coefficients above the diagonal, corrected below.

Since the group is very homogeneous in respect of age, the correlations with age are all low, and the effect of partialing out age variability very slight. Nevertheless, all subsequent computations are based upon the partial correlations in Table V, in order that even the slight correction for age may be utilized.

All the correlation coefficients have been reported to four decimal places. This must not, of course, be interpreted to mean that they are definitely determined to the 4th decimal. The accuracy[1] with which a correlation coefficient has been determined can best be indicated by its probable error. The P.E. of the largest raw r in Table V, r_{12}, is .0328, that of the smallest, r_{17}, is .0570. The P.E.'s of the other correlations will lie between these two extreme values. This shows that the correlations are barely established to one decimal place, and can deviate considerably from the "true" value even in the second decimal place. The main reason for reporting the correlations to four places is that four place correlations must be used in subsequent computations, such as tetrads which frequently yield very small values. It has been considered preferable to state the actual values which will be used in these computations, rather than the nearest two-place values, in order to furnish the necessary data for anyone wishing to check the computations.

The correlations in Table V suggest very strongly the presence of three fairly independent group factors, running through the tests. The correlations within the memory, or verbal, or numerical groups, for instance, are substantially higher than the correlations between any two groups, as an examination of Table V will reveal. To bring this clustering out more clearly, average correlations within each group have been computed. This was done by first transmuting each correlation coefficient into z-scale units (cf. 7, p. 163 to 171), averaging these z's, and then transmuting this average value back into a corresponding correlation coefficient. This technique is preferable to the customary procedure of averaging correlations directly, since successive z-units represent equal steps, whereas successive r-units do not. Correlation units at the upper end of the scale correspond to much larger differences in degree of relationship than those at the lower end of the scale, hence it is especially necessary to use the z function when the correlations are numerically high, as they are in

[1] We here use the term "accuracy" to refer not to computation errors, but to the closeness with which our data approximate the "true" values which would be obtained if the entire population were tested, rather than a sampling of it. The argument offered above to justify our reporting the data to four decimal places is one commonly given in statistical work. The reader is referred, for example, to Pearson's justification of such a procedure in connection with data reported by him in 1905: (21), p. 53-54.

the present data. The average correlations, computed by this method, are given in Table VI.

TABLE VI
AVERAGE CORRELATIONS

Correlations Averaged	*Average of Raw Correlations*	*Average of Correlations Corrected for Attenuation*
6 intercorrelations of 4 memory tests	.4136	.5580
8 correlations: each memory test paired with each verbal test	.0997	.1391
8 correlations: each memory test paired with each numerical test	.1586	.1974
Single correlation between the two verbal tests	.6513	.8153
Single correlation between the two numerical tests	.5805	.6390
4 correlations: each verbal test paired with each number test	.1586	.1974

With a sample of 140 cases a correlation coefficient must be at least .22 in order to be significant, i.e., over four times as large as its P.E. The average correlations in Table VI seem to substantiate quite strongly the earlier results on the presence and mutual independence of the three factors of memory, verbal ability, and numerical ability.

Some of the individual correlations are of interest. The highest correlation, for instance, between a single memory test and a single non-memory test is that between syllable recognition and analogies ($r = .3005$). This might be expected, however, since the analogies test was a multiple choice recognition test. The vocabulary test also correlates higher with syllable recognition than with any of the other memory tests ($r = .1490$). Among the memory tests themselves, the highest correlation is between word-word and retained members ($r = .6078$). This correlation is much higher than that between word-word and picture-number (.5090), suggesting that community of content is probably more potent than community of method in determining the correlation between two tests. It can be shown, however, that the correlation between word-word and retained members is not attributable, even in part, to "verbal ability" *as measured by the two tests* of it which we used. The correlation between word-word and retained members, with variability arising from vocabulary held

constant by partial correlation, is .6086; partialing out variability due to both vocabulary and analogies simultaneously, the correlation rises to .6150. This result does not, indeed, negate entirely the interpretation of the obtained correlation in terms of community of content, as the present "verbal" tests hardly measure *all* aspects of verbal ability.

III. *The Tetrad Analysis*

In order to demonstrate more fully the presence and independence of the memory factor, tetrad equations have been computed for various combinations of our tests. These test combinations include, first, the four memory tests; secondly, every possible pair of memory tests combined with two verbal tests; thirdly, every possible pair of memory tests combined with two numerical tests; and finally, the two verbal and the two numerical tests taken together. The values of the tetrad differences obtained are given in Table VII.

TABLE VII
TETRAD DIFFERENCES*

Variables	t_{1234}	t_{1243}	t_{1342}
Memory Tests:			
5678	−.0130	−.0196	−.0066
	±.0316	±.0257	±.0312
Verbal and Memory:			
1256	.3302	.3290	−.0112
1257	.3953	.3991	.0038
1258	.2048	.1850	−.0198
1267	.3024	.3047	.0023
1268	.1367	.1499	.0132
1278	.1753	.1669	−.0084
Numerical and Memory:			
3456	.2804	.2805	.0001
3457	.3287	.3335	.0048
3458	.1668	.1856	.0188
3467	.2312	.2390	.0078
3468	.1015	.1324	.0309
3478	.1028	.1402	.0374
Verbal and Numerical:			
1234	.3407	.3206	−.0201
	±.0322	±.0333	±.0121

*Key to variables:
1—Vocabulary
2—Analogies
3—Number series completion
4—Arithmetic reasoning
5—Word-word
6—Picture-number
7—Retained members
8—Syllable recognition

Each row in Table VII contains the three tetrads computed from the four variables indicated in the first column. All of the tetrad differences have been computed from the raw correlation coefficients of Table VI, with age variability constant. The P.E.'s have been computed[2] for each of the three tetrad differences involving the four memory tests, and are given in Table VII. None of these tetrad differences is as large as its P.E., the ratios $T.D/PE_t$ being .4, .8, and .2 respectively. The chances in 100 that these values indicate "true" differences greater than zero are very low, being 11, 19, and 5, respectively. It seems clear, therefore, that these four variables satisfy the tetrad criterion, indicating that a single common factor is adequate to account for their intercorrelations. This result corroborates our 1930 study.

Turning now to the tetrads in which pairs of tests of each trait have been combined, we find equally striking evidence for the independence of this memory factor from the other traits measured. These tetrads all conform very closely to the type described by Kelley (14, proposition No. 16, p. 69). In that proposition, Kelley demonstrates that, "If the intercorrelations between four variables are such that $t_{1234} = t_{1243}$ and $t_{1342} = 0$, they could conceivably have arisen from four variables x_1, x_2, x_3, and x_4 through which was a general factor plus, in addition thereto, a second factor common to x_1 and x_2, or a second factor common to x_3 and x_4." The same type of tetrad differences results if there is a group factor through x_1 and x_2, and another group factor through x_3 and x_4. When any two of our memory tests are combined with the two verbal or the two numerical tests, the first two tetrad equations in the set give very large and significant deviations from zero, and are equal to each other within their P.E.'s; the third tetrad difference is never a significant deviation from zero. The results are so clear cut and consistent that it hardly seems necessary to report the P.E.'s of the tetrads in order to estab-

---

[2] The formula used is that reported by Kelley (14), p. 49:

$$P.E._{t_{1234}} = \frac{.6745}{\sqrt{N}} \Big[r^2_{12} + r^2_{13} + r^2_{24} + r^2_{34} + 2r_{12}r_{14}r_{23}r_{34} + 2r_{13}r_{14}r_{23}r_{24} - 2r_{12}r_{13}r_{23} - 2r_{12}r_{14}r_{24} - 2r_{13}r_{14}r_{34} - 2r_{23}r_{24}r_{34} + t^2_{1234}\,(r^2_{12} + r^2_{13} + r^2_{14} + r^2_{23} + r^2_{24} + r^2_{34} - 4) \Big]^{1/2}$$

lish this fact. With 140 cases, a tetrad difference of the order of .2 or .3 is obviously a significant deviation from zero. However, for the sake of greater certainty, P.E.'s have been computed for the largest and smallest tetrads in each set in order to give some idea of the range between which the P.E.'s fall. These P.E.'s have all been computed by the formula given by Kelley (14) and are reported in Table VIII.

TABLE VIII

PROBABLE ERRORS OF SELECTED TETRADS

Variables	*Designation of Tetrad*	*Numerical Value of Tetrad Difference*	$P.E._{t}$	$t/P.E._{t}$
Verbal and Memory:	t_{1672}	.0023	.0035	.6571
	t_{1582}	.0198	.0171	1.1579
	t_{1275}	.3991	.0320	12.4719
Numerical and Memory:	t_{3564}	.0001	.0072	.0139
	t_{3784}	.0374	.0146	2.5616
	t_{3475}	.3335	.0314	10.6210

If we consider first the tetrads involving the verbal and memory tests, the P.E.'s given in Table VIII are those of the smallest and largest of the tetrad differences which should, theoretically, equal zero, i.e., the third tetrad in each set of three. The P.E.'s of these tetrads are .0035 and .0171, respectively. Both tetrad differences are much smaller than four P.E. Of the tetrads which should theoretically be significantly greater than zero, $t_{1275} = .3991$ is the largest. The P.E. of this tetrad is .0320, making the tetrad over 12 times as large as its P.E. and therefore significant. The tetrads and P.E.'s given in Table VIII for the numerical and memory tests have been selected in the same way, and they also support the theoretical expectations.

The tetrad equations with the verbal and numerical tests combined (Table VII), again conform to Kelley's 16th proposition, indicating the possibility of a group factor through the two verbal tests and another group factor through the two numerical tests. The first tetrad difference is 10.58 times as large as its P.E., the second 9.63 times as large, and the third 1.66. Again we find that t_{1234} and t_{1243} are large and significant deviations from zero and t_{1342} is not significantly greater than zero. These results agree very closely with those

of Schneck (26) as regards the verbal and numerical factors. This is shown by the following comparison:

	Tetrad Differences Obtained in Schneck's Study	*Tetrad Differences Obtained in the Present Study*
t_{1234}	.3120	.3407
t_{1243}	.3132	.3206
t_{1342}	.0012	—.0201

IV. *Correlations with Central Factor and Regression Equation*

The correlations of each of the four memory tests with what is common to all four were computed by Spearman's formula (29 p. XVI, Appendix) which may be expressed as follows, when *four* tests are used:

$$r_{1g} = \sqrt{\frac{r_{12}\, r_{13}}{r_{23}}} = \sqrt{\frac{r_{12}\, r_{14}}{r_{24}}} = \sqrt{\frac{r_{13}\, r_{14}}{r_{34}}}$$

Since the tetrad differences with the four memory tests were not equal exactly to zero, the three determinations of r_{1g} are not identical. In such a case, the customary procedure is to compute all of the three determinations and then to average them in order to arrive at the best approximation. This procedure has been followed in the present study, except that the correlations were first transmuted into z-scale units and then averaged (cf. p. 16). The following are the average correlations obtained by this method:

word-word:	$r_{5g} = .8426$
picture-number:	$r_{6g} = .6291$
retained members:	$r_{7g} = .7211$
syllable recognition:	$r_{8g} = .3800$

With these correlations, a regression equation was computed, from which scores in the central memory factor could be estimated. The equation is expressed in terms of reduced scores, the S.D. of each distribution being taken equal to 1.00. In the equation which follows, x_o represents the central mem-

ory factor and x_5, x_6, x_7, and x_8, respectively the four memory tests, following the numbering used throughout this study.[3]

$$x_0 = .5467\, x_5 + .2042\, x_6 + .2729\, x_7 + .0806\, x_8$$

It will be seen that the word-word test has by far the largest weight in determining the central factor, the retained members test ranks second and the picture-number third, the syllable recognition test having very little weight. Apparently the central factor is largely verbal memory and is best measured by the word-word paired associates test. The multiple correlations of successive groupings of the four tests with the central factor further support this conclusion. The correlations which follow were computed by the usual multiple correlation formula (8 p. 262, formula No. 58), making use of the r_{ag}'s as well as the intercorrelations of the four memory tests:

$$\begin{aligned} r_{05} &= .8426 \\ R_{0(56)} &= .8740 \\ R_{0(567)} &= .9001 \\ R_{0(5678)} &= .9032 \end{aligned}$$

The correlation of the entire battery with the central factor is very high, .9032. This indicates that the battery measures the common factor found, whatever it may be, very fully. Or, stated conversely, it suggests that the common factor found is defined quite completely by the nature of these particular tests, and contains little else besides what is found in these tests. We shall have occasion to refer to this point later.

---

[3] The formula used was:

$$\begin{aligned} x_0 = {} & r_{01.234} \frac{\sqrt{1 - r^2_{04}}\ \sqrt{1 - r^2_{02.4}}\ \sqrt{1 - r^2_{03.24}}}{\sqrt{1 - r^2_{14}}\ \sqrt{1 - r^2_{12.4}}\ \sqrt{1 - r^2_{13.24}}}\, x_1 \\ & + r_{02.134} \frac{\sqrt{1 - r^2_{01}}\ \sqrt{1 - r^2_{03.1}}\ \sqrt{1 - r^2_{04.13}}}{\sqrt{1 - r^2_{12}}\ \sqrt{1 - r^2_{23.1}}\ \sqrt{1 - r^2_{24.13}}}\, x_2 \\ & + r_{03.124} \frac{\sqrt{1 - r^2_{01}}\ \sqrt{1 - r^2_{02.4}}\ \sqrt{1 - r^2_{01.24}}}{\sqrt{1 - r^2_{34}}\ \sqrt{1 - r^2_{23.4}}\ \sqrt{1 - r^2_{13.24}}}\, x_3 \\ & + r_{04.123} \frac{\sqrt{1 - r^2_{01}}\ \sqrt{1 - r^2_{03.1}}\ \sqrt{1 - r^2_{02.13}}}{\sqrt{1 - r^2_{34}}\ \sqrt{1 - r^2_{14.3}}\ \sqrt{1 - r^2_{24.13}}}\, x_4 \end{aligned}$$

This formula is so expressed that the third order partials in numerator and denominator cancel throughout, thus saving considerable labor in computation.

V. *Analysis of Variance*

From the correlation of each test with the central factor, it is possible to find the degree to which individual differences in performance on each test are determined by differences in the common factor, i.e., the percentage of the *variance* of each test attributable to the variance of the common factor. It can be readily demonstrated that this is simply the square of the r_{ag} of each test (cf. 2, p. 53-54). The remainder of the variance is attributable to factors specific to each test, *plus* errors of measurement. The variance attributable to all true ability factors (both general and specific) is equal to the reliability coefficient of the test,[4] the difference between this value and the value for the general factor alone giving the contribution of the specifics alone, freed from errors of measurement. The percentage of the variance of each of the memory tests attributable to each of the three classes of factors is given in Table IX.

Table IX

PERCENTAGE OF VARIANCE ATTRIBUTABLE TO COMMON AND SPECIFIC FACTORS, AND TO ERRORS OF MEASUREMENT

	% of Variance Attributable to		
Test	*Common Factor*	*Specifics*	*Errors of Measurement*
5. Word-Word	71	15	14
6. Picture-Number	40	43	17
7. Retained Members	52	14	34
8. Syllable Recognition	14	50	36

This analysis of variance suggests some interesting possibilities regarding the relation of the different tests to the central factor. The word-word test is based very largely on the central factor, with a very small contribution from specific factors and errors of measurement. Over one half of the variance of the retained members test is determined by the central factor, *even though as much as 34% is attributable to errors of measurement.* Contrast this with the picture-number test, in which, with only 17% of the variance attributable to chance

[4] The index of reliability of a test, $\sqrt{r_{II}}$, gives the correlation of the test with a perfect measure of the abilities involved. Hence $(\sqrt{r_{II}})^2$, or r_{II}, gives the % of the variance of the test attributable to the variance in the true abilities involved, freed from errors of measurement.

errors, the percentage due to the common factor is only 40. In other words, in the retained members test, it is chiefly chance errors which keep the correlation with the central factor from being higher, whereas in the picture-number test the correlation is no higher because of the large contribution of specifics in that test. Syllable recognition also has prominent specific factors, in addition to large errors of measurement. We again suggest, therefore, that the common factor is to be defined largely in terms of such tests as the word-word and retained members, i.e., memory for a certain *type* of *material*, rather than memory tested by some one *method*.

VI. *Conclusions*

1. The first conclusion to be drawn from the results of this study is that they fully corroborate the former results on the memory, verbal and numerical factors, and their mutual independence. The very close agreement between the earlier results and those secured at present lends a certain stability and positiveness to the data. Ordinarily, if certain data have low probable errors, statistically estimated, they are regarded as reliable. The present study may be considered, from one angle, as an experimental verification of such reliability. The obtained data have proved to be in nearly every case even more "reliable," in this sense, than the probable errors would have led us to expect.

2. In the second place, our population has been extended, and may now be defined as including college students of either sex, very probably from any of the four years of college and anywhere within the age range of college students, but of a specified racial and cultural background. (Cf. description of subjects in Ch. II, p. 9 to 10.) The latter limitation should still be included, because of the very likely influence of general training as well as formal education on mental organization.

3. Finally, the data have offered some—very tentative—suggestions regarding the possible nature of the common factor found through the memory tests. The fact that the common factor is measured so completely by the particular battery of four tests used suggests either of two possibilities. One is, that our tests measure "memory" in general very extensively and thoroughly—this is rather difficult to expect, since the tests measure a very simple and narrow form of memory.

Another possibility is that the common factor found is itself very limited in scope, and that its nature is defined quite completely by the type of activity involved in *these tests only*. The argument followed in this discussion is that if our battery had measured the central factor *less completely* than it does, if, in other words, the multiple correlation between the four tests and the common factor had been *lower*, we should more readily expect the common factor to extend beyond these tests to other types of memory. A further suggestion regarding the nature of the common factor is that community of content[5] seems to be more potent than community of method in determining inter-test correlations. We shall return to these suggestions in the course of Parts II and III of this monograph, where their implications will be followed through more thoroughly.

[5] In the sense of actual test material, not mental content.

PART II*

Chapter IV

PROCEDURE

I. *The Subjects*

The subjects in this study consisted of women students at Brooklyn College enrolled during the Spring semester in the same courses from which the subjects in Part I had been drawn during the Fall semester.[1] The number of subjects originally tested was 217. Of these, 47 were eliminated because of absence or other irregularities in the records, leaving a total of 170 used in the final results. Each subject filled out a questionnaire, just as in Part I. The results of this questionnaire, given in Table X, show that this sampling was very similar to that used in Part I, in every respect. The data in Table X are expressed in the same terms as in Part I, all the figures except those in the "age" columns being percentages.

II. *The Tests*

A total of ten tests were given to all the subjects in this study. Of these ten, two were selected from the memory tests used in the earlier study, one was a revision of a "verbal" test used by Schneck (26), two were standardized tests in general use, and the remaining five were especially constructed for this study. It was essential to include a non-memory test in our set-up, for otherwise, if a common factor should be found through all the memory tests, we could not tell whether it was a *memory* factor, or a general ability factor operative in all mental tests of this sort. If a memory factor is present in our tests, its independence can be tested out by correlating the memory tests with the non-memory test. For this purpose, the analogies test used in Part I was chosen, firstly because it

* The writer is indebted to Mr. G. W. Lawler of Brooklyn College for his cooperation in obtaining the subjects and for administering the tests in this part of the study.

[1] Since it was shown in Part I that the results obtained on this population coincided very closely with the results of the 1930 study, we may assume conversely that the findings in Part II will apply also to the population of the earlier study.

Table X

ANALYSIS OF SUBJECTS

Class		*Age*	*Birthplace*		*Father's Birthplace*	*Mother's Birthplace*
Low Junior	2	Mean = 18.42	United States	96	19.5	31.5
Sophomore	95.5	S.D. = 0.93	Russia and Poland	2.5	51	48.5
Upper Freshman	2.5	Range: 16.58 to 21.42	Austria	0	18	12.5
		Sk/σ_{Sk} = —0.63	Germany	0	2.5	1
		Ku = .3029 ± .0705	Italy	1	5	3
			Norway and Sweden	0	0.5	0
			British Isles	0.5	2	1
			Roumania	0	1.5	2.5

Jewish		*Other Language Spoken at Home*		*Father's Education*		*Father's Occupation*		*Hours of Outside Employment per Week*	
Yes	86	Yes	58.5	None in U. S.	24	None Specified*	10	None	70.5
No	14	No	41.5	Elementary School	33.5	Skilled Labor	20	1– 4	2
				High School	28	Merchants, Shopkeepers, and Clothiers	39	5– 9	20
				College and Professional School	14.5	Clerical and Selling	19.5	10–14	3.5
						Professional and Semi-Professional	11.5	15–19	1
								20 and Over	3

*Includes: deceased, retired, and unemployed.

is one of the best single measures of Schneck's verbal factor,[2] and secondly because tests of analogies such as this have been repeatedly held up by the Spearman School as being among the best single measures of *g*. In this study, a revision of Schneck's original test, which had proved to be somewhat better suited to college students, was adopted. This revision contains 78 items, instead of the original 40, and is given with a time limit of 30 minutes.

Two of the original tests of immediate memory for visually presented rote material were included as reference values, so that if a common factor should be found through the new memory tests, we could determine whether it is the same as the memory factor previously found. For this purpose, the word-word and syllable recognition tests were chosen because they differ in method and content and yet both had been shown to measure the common factor found in the earlier study. The syllable recognition test was used exactly as it had been in the previous study. The word-word test differed from the original only in having a fore-exercise, in which the subjects were shown five pairs of words not used in the test proper and were tested on these pairs in the same manner as in the main test. This modification was introduced in order to facilitate the understanding of directions, so that no papers would have to be discarded on this account. Moreover it was hoped that we might reduce to a minimum the influence upon test scores of individual differences in the understanding of directions. Insofar as possible, the fore-exercise was used in all the new memory tests, for the same reasons.

Our new tests of memory differ as much as possible from each other. Since the chief purpose of this study was to discover if the common factor previously found through a selected group of memory tests was general retentivity, the attempt was made to employ tests which differ in every possible respect except that of retentivity. That is, our tests all involve the retention of impressions over varying time-intervals, but the specific materials as well as the methods of presentation and of testing differ. The interval of delay varied from practically zero to two days. In some tests, retention was tested immediately after the presentation of each *item*; in others, retention was tested at the end of the entire series

[2] Its correlation with the central verbal factor was .82 (26).

of items; and in one test of "delayed" memory in the common sense of the term, the interval between presentation and testing was two days. The specific tests used are: logical recall, logical recognition, delayed memory for words, incidental recognition of forms, memory for sequences of movement, and the Seashore test of memory for tonal sequences. Since the last test has frequently[3] been found to correlate very highly with ability to discriminate the pitch of the sounds heard, the Seashore pitch discrimination test was also given, in order to find the relative contribution of retentivity and pitch discrimination to performance in the tonal memory test. The new tests will be described more fully in the following section.

III. *Construction of New Tests and Preliminary Experimentation*

In constructing the two logical memory tests, the attempt was made to reduce individual differences in language comprehension to a minimum. Sentences were short and simply constructed, and the passages were written so as to be well within the comprehension of college students in general. The passages were descriptive and narrative, rather than expository. They were rich in concrete detail and did not require the subjects to grasp general principles or abstract ideas. In order to reduce individual differences arising from past training or special interest in a particular field, the use of technical terms was avoided; the material used in recall and in recognition was drawn from different fields; and finally, the passages were chosen so as to be about equally unfamiliar to all the subjects.

The material finally chosen as best suited to the purpose consisted of two prose passages, one describing the architecture and building customs of a fictitious primitive race, and the other dealing with some of the less familiar aspects of protective coloration in animals. Both passages were read aloud by the experimenter to the subjects. Auditory presentation was decided upon because all of the memory tests used in the former study had been visual, and it was deemed expedient to study memory through another sense modality. Moreover, from the experimental point of view, the presentation of a prose passage is better controlled and better standardized if

[3] Cf. (5) and (16).

the experimenter reads it. If the subjects read it, the quicker readers will probably finish the passage and reread it, when a constant reading time is allowed; and if they are directed to read it through only *once*, even assuming complete cooperation, some subjects will spend less time on the passage than others. In either case, uncontrolled factors would be introduced, which can be eliminated by a properly standardized auditory presentation. Since these tests were to measure logical memory, the passages were read at a uniform rate but with natural voice modulations, arranged so as to bring out the meaning. This was considered preferable to a perfectly monotonous presentation. The rest pauses, as well as the words or phrases to be emphasized, were indicated by appropriate symbols on the copy. The reading of each passage required approximately 12 minutes.

The attempt was made to have the recall and recognition tests as comparable as possible, as the results were to be used also in a study of recall and recognition reported in Part III. Accordingly, the two passages were of the same length, the directions very similar, and the number of items in the two tests identical. Recognition was tested by a five-alternative multiple choice test, recall by a series of very specific short-answer questions, in which the answer was to be a word or phrase. The items appeared in the test series in the same sequence as in the passages. In the original form of the tests, there were 100 items in each. After the test had been given to a group of 20 male college students in a preliminary tryout, several of the items were revised and the tests shortened to 80 items. This reduction was made necessary because it was found difficult for all of the subjects to finish the test within the 50-minute period that would be available in the experiment proper. Furthermore, the original tests proved somewhat too difficult, yielding skewed distributions.

In constructing the incidental memory test, our main concern was to get a test situation in which the materials would be carefully attended to and observed, but with no attempt on the subjects' part to memorize them. The test finally chosen was a form recognition test consisting of two parts. Part I was a test in the estimation of areas. A series of 30 two-dimensional geometrical figures drawn in India ink on white cards, were presented one at a time. A series of 10 squares,

similarly drawn, was visible continuously to the subjects. Each of the 30 figures shown had been drawn so as to be exactly equal in area to one of the 10 squares. The subjects were instructed to estimate the area of each figure by indicating the proper square in each case. During this part of the test, the subjects were given no indication of the fact that they were to remember anything that was being shown. Immediately after the completion of the estimation of areas test, a recognition test was given in which the original 30 figures were mixed in a random but predetermined order with 30 new figures. This test was tried out on 18 women students in experimental psychology at Barnard College, with the special purpose of determining number of items to be included and questions of *timing.* It was not found necessary to introduce any modification in the test. The timing finally decided upon was 5″ exposure with 3″ interval for each card in Part I, and 3″ exposure with 2″ interval for Part II.

The delayed memory test was a retained members test for words, the words having been selected from the original retained members test which had been used as a test of immediate memory. A list of 30 out of the original 40 words was tried out on 20 male college students. Since the test proved to be too difficult, the median score being only 5.5 words correct, the test was further cut down to only 20 items. The delay chosen was exactly two days, the two parts of the test being given on two successive meetings of the class, at the same hour of the day. Presentation of the words was visual, as in the original test.

The "memory for movement" test was based in principle on the "Imitation of Movement" test in the Pintner non-language scale (23). The test is administered very similarly to a memory span test, since testing is *immediate,* after each series of movements, and the response must be completely correct in order to count at all. The movements were made with a pointer to and from each of four dots. The dots were painted in black India ink on a sheet of white cardboard which was tacked on the wall in front of the subjects. This was considered preferable to using chalk dots on the blackboard, since when using chalk, the pointer leaves streaks between the dots which may either aid or confuse the subjects in recalling the movements made. The experimenter moved the pointer from dot to dot

according to certain specified sequences, and at the end of each series of movements, the subjects reproduced them in pencil on specially prepared mimeographed sheets. The length of each series varied from four to seven movements. Movements were made both from left to right and from right to left, the former always being above the dots, the latter below. Any one dot could be touched repeatedly by the pointer during a single series, and not all the four dots were used in each series. The movements were made at the rate of one per second. A two-minute rest period was given after the twelfth and thirty-first series, which corresponded to the end of the first and second pages of the subjects' data sheets. This was considered necessary since the test otherwise causes considerable strain and tension. The test was tried out on 30 women students in introductory psychology at Barnard College. The number of series finally decided upon for the test was 50. These series were arranged in an approximate order of difficulty from easiest to hardest, on the basis of the preliminary results on the Barnard group. It was decided to suggest to the subjects, in the directions, the common device of mentally numbering the dots and counting while the movements are being made, so as to keep track of the sequence. Several subjects usually hit upon this device in the course of the test anyway, which would introduce a spurious condition from our standpoint, since we want to measure individual differences in *retention,* not in the ingenuity with which the subjects attack the test. By telling all the subjects about this device beforehand, the test is made more standard for everyone.

IV. *Testing Procedure*

All of the tests were given during lecture periods, to sections of from 15 to 40 students. The tests were given by two examiners, the regular instructors in the classes used. Those tests which required more elaborate control of timing, or demonstrations by the examiner, were all given by one of the two examiners, in order further to standardize the procedure. Very detailed typewritten directions were given to each examiner, describing the exact procedure to be followed and also giving certain oral directions and explanations to be read by the examiner to the subjects. The examiner did not say anything during the tests except what was on the typewritten

sheets. In addition, each subject had further written directions on the individual data sheets. Special mimeographed data sheets were used for each test except the two Seashore tests for which the standard printed blanks furnished by Stoelting & Co. were used. The specific directions read to the subjects in *each test* follow. These directions are reproduced in full since the directions used so frequently establish a mental set and affect materially the nature of the test itself.

Analogies

The directions for this test were mimeographed on the data sheets. The examiner reads the directions aloud while the subjects follow them on their sheets. The directions were identical to those used by Schneck in the earlier form of this test (cf. 26, p. 16).

Memory for Movement

"This is a test of your memory for a series of movements. I will make the movements with a pointer on these four dots (point). As soon as I finish each series of movements, you are to reproduce the movements in pencil, using the rows of four stars on your sheets as I do these dots. The movements must all be in the correct order and correct direction." (Examiner gives the three sample series. Subjects reproduce each series, and examiner indicates the correct responses on the blackboard afterwards. Examiner counts aloud while drawing in the movements.) "It will help you to count to yourself while the movement is being demonstrated on the chart, so that you can keep track of the sequence of the dots. Do not make any marks until I am through with each series of movements. Be sure to make your drawings on the proper row of stars each time. I shall give you the number of the series in each case, so you can keep track of them. There will be 50 series altogether. Ready. . . ."

Word-Word Test (Exposure 3″—interval 2″ for each item)

"This is a test of memory by the method of paired associates, that is, you will be shown a series of cards each containing a pair of four-letter words. When all the cards have been shown, they will be shuffled and you will be shown only the *first* of each pair of words; you are to write down the second word that went with it in each case. We will go through a sample set of 5 cards first, for practice. Observe the cards carefully, but write nothing while they are being shown." (Examiner presents the 5 demonstration cards, then rearranges them for the test series.) "Now turn over the data sheets. As each word is shown, write the other word that went with it. Write your answers in the column headed 'sample.'" (Examiner reads off the correct responses after completion of the sample series.) "Now I will show you 20 cards, similar to those in the sample set. Observe them in the same way and write nothing until told." (Present cards.) Now turn to the data sheets again. As each word is shown, write the word that belongs with it, in the proper place under Part I. If you cannot remember a word, draw a line in its place, but do not write the next word in its space. Each word must be next to the right number on the sheet in order to be scored correctly. I will read you the number of each card as I hold it up, so that you can keep track of them. Ready. (Procedure in Part II, with the second set of 20 cards is identical to Part I.)

Nonsense Syllable Recognition (Exposure 3″—interval 2″ for each item)

"I am going to show you a series of 20 nonsense syllables, that is, combinations of letters having no meaning in the English language. I

will then show you a series of 40 syllables, 20 of which are the same as those in the first list, and the other 20 are new, and you are to try to recognize the old syllables. Observe carefully, but write nothing while the first series is shown. Ready. . . ." (Examiner presents cards.) "Now I will show the second series. As each syllable is shown, you are to write a plus sign if you have seen it before, and a minus sign if it is a new syllable. Do not skip any spaces or make any omissions. If you cannot remember whether or not you have seen a syllable before, guess. I will read the number of each syllable as I hold up the card so that you can keep track of them. Ready. . . ." (The procedure in Part II, with the second set of 40 syllables is identical to Part I.)

Logical Recognition and Logical Recall

General directions read to the subjects before presenting each passage: "This is a test of memory for a connected passage. Listen carefully while I read the passage and try to memorize all the details. As soon as I finish reading the passage you will be tested for recognition (recall) of every specific item of information in the passage." After reading the passage, data sheets are passed out, on which further directions are mimeographed. Examiner reads these directions aloud while subjects follow them on their sheets. The specific directions for recognition and for recall are as follows:

Logical Recognition

The following is a series of statements referring to the passage that has just been read. Each statement is followed by *five alternatives*. You are to write (in the space provided to the right of each statement) the *number* of the correct alternative in each case. Base your decision *only* on the passage just read, disregarding any other information you may have from other sources. If some of the statements refer to facts not directly discussed in the passage, or if none of the given alternatives is correct, then write the number "5" which means "omitted" and indicates that the question cannot be answered correctly from the information given in the passage.

The following are examples:

(a) The chief subject discussed in this passage is: (1) plant metabolism; (2) animal coloration; (3) heredity; (4) American bird life; (5) omitted. ...2...

The correct answer here is "animal coloration," hence we write the number 2 in the answer space.

(b) The spotted coloration of the giraffe is classed as: (1) disruptive; (2) mimicry; (3) protective; (4) warning; (5) omitted....5...

Here "5" is the correct answer, since the giraffe was not discussed in the passage read. Even though you may know from other sources, or *infer* from the passage just read, that the correct answer is one of the other alternatives, "5" should be the answer given.

Proceed in exactly the same way with each of the statements below, taking them in order. Write only one number in each space.

Logical Recall

The following is a series of questions referring to the passage that has just been read. Write the answer to each question on the dotted line beneath it. Base your answer *only* on the passage read, disregarding any other information that you may have from other sources. This is a test of *memory, not knowledge!*

Your answers should consist, in most cases, of one or two words or a phrase. *Do not explain* the point, but try to remember what was said about it in the passage, and write that down.

The following are examples:

(a) What period in the history of the Dekka tribes is referred to in the passage read?

..........................period of conquest..........................

(b) In the case of a people inclined to building, what exerts a profound influence on the results achieved?

................nature of building materials available................

In the second question, several other factors would be correct, on the basis of general information. But the answer given—based solely on the passage read—is the *only* correct answer from the point of view of this memory test.

Proceed with the questions below, taking each in order. Try to put down an answer to each question, even if you are not sure.

Delayed Memory for Words (Exposure 3″—interval 2″ for each item)

"This is a test of delayed memory. You will be shown 20 four-letter words, one at a time. At the next meeting, you will be asked to write down as many of these words as you can remember, *in any order*. Try to memorize the words while they are being shown, but do not rehearse any of the words afterwards, nor talk about them. Try to forget all about the test, until next time. Ready. . . ."

Directions for second period, after a 48-hour interval:

"Write down as many of the 20 words shown last time as you can remember. Write the words in the order in which you remember them, putting each down as you think of it." (Allow 10 minutes for recall.)

Incidental Memory Test

Part I: Estimation of Areas (Exposure 5″—interval 3″ for each item)

"This is a test of your ability to estimate the areas of various figures. You will be shown a series of 30 figures, one at a time. Each figure is exactly equal in area to one of the 10 squares on the board, although the figures have different shapes, some are squares, others triangles, circles, etc. As each figure is shown, you are to write down the letter corresponding to the square which you think has the same area. You must write a letter for each figure shown; if you are not sure, guess.

In judging the areas, consider the *entire* area enclosed within the *outer* edges of the figure, not just the white space. For instance, this triangle (hold up Demonstration Card No. 1) has the same area as this one (hold up No. 2), in spite of the wide border. This figure (hold up No. 3), on the other hand, has a smaller area, since it is not a complete triangle. Consider only the area within the edges of the figure, without filling in. Some of the figures will have designs on them, like this (hold up No. 4). The design does not alter the area of the figure. In this case, you should judge the area covered by the entire square. What letter would you write in giving the area of this figure? (Get oral answers from subjects.) "C" is the right answer. You are to do the same for each figure shown. I will read the number of each card, from 1 to 30, as it is shown, so that you can put the answers next to the right numbers on the sheet. Make no omissions. Be sure to write one of the 10 letters in each case. Ready. . . ."

Part II: Recognition Test (Exposure 3″—interval 2″ for each item)

"Although you were not told to try to remember the figures shown, we want to see how well you *can* remember them. You will be shown the original 30 figures mixed with 30 new ones. See if you can recognize the old ones. As each card is shown, write a plus sign if you have seen it before, and a minus sign if it is a new figure. Consider every detail in the figure in deciding; if you are not sure, guess. Ready. . . ."

Seashore Tests

The Tonal Memory and the Pitch Discrimination tests were given with the standard directions reported by Seashore in his "Manual of Instructions and Interpretations for Measures of Musical Talent." (25)

The testing extended over a period of four months, occupied a total of approximately five hours and ten minutes, and was scattered over nine class periods. The sequence in which the tests were given, together with the number of items and the approximate duration of each test are given in Table XI.

TABLE XI

TESTING SCHEDULE

Period	*Tests, in the Order Given*	*Number of Items*	*Approximate Duration of Entire Test in Minutes*
I	Analogies	78	35
II	Memory for Movement	50	20
III	Word-Word	40	15
	Syllable Recognition	80	20
IV	Logical Recognition	80	50
V	Logical Recall	80	50
VI and VII	Delayed Memory	20	5, 10
VIII	Incidental Memory	60	20
IX	Tonal Memory	50	25
	Pitch Discrimination	100	25

It should be added that the incidental memory test was given at the time when the students were on the topic of "perception" in class. It seemed quite natural to the students, at such a time, to be given a test on the estimation of area, in which certain illusions of size were to be demonstrated. None suspected that it was to be a memory test, since the discussion of memory in class had been completed several weeks before.

V. *Scoring*

The score in each test is in terms of total number of items correct; no partial credits were given in any test. The maximum score possible in each case is therefore equal to the total number of items in the test. A word of explanation should be added regarding the scoring of the recognition tests. The score in the nonsense syllable recognition test was taken as the total number of items minus *twice* the number of wrong items, thus allowing for guessing. This scoring method was used, since it had been the method followed with this test in the two previous studies and it was desired to express the data in comparable terms. In the incidental recognition test, however, the

score is simply the total number of items correct. This simpler scoring method could have been used in both tests, for the purposes of the present study, since it can be readily shown (12) that when no omissions are made, as was the case in these two recognition tests, a perfect correlation exists between the "number right" score, and the "total minus twice the number wrong" score. No correction for guessing was used in the logical recognition test, since there are five alternative answers in each item. The chances of guessing correctly in this test are no greater than in tests such as Analogies and Vocabulary, in which no correction for guessing is ordinarily made. Guessing was further reduced in the logical recognition test by allowing the subjects to omit items rather than requiring them to write an answer for each item as in the other two recognition tests. The use of raw scores is especially permissible in the present study, since our interest here was not in the actual *amount* correct, but in inter-test correlation.

On the two Seashore tests, tonal memory and pitch discrimination, raw scores were used rather than the percentage scores used by Seashore in his tables of norms. Here again, either method of scoring can be used interchangeably for correlation purposes.

Chapter V

RESULTS

I. *Evaluation of the Tests*

In Table XII are presented the data on reliability, variability, and normality of distribution of each of our ten tests. Six of these tests have reliability coefficients that are well over .80. The Analogies test has a reliability of .70 which corroborates very closely the reliability coefficient found with a similar form of the test in Part I (cf. p. 13). Delayed memory has a reliability of only .64, but considering the shortness of the test, this reliability is probably satisfactory. The reliability coefficient of syllable recognition is rather surprising, being .5256 ($\pm$ .0374), as compared to .6369 ($\pm$ .0307) obtained in Part I with the same test. This difference may conceivably have arisen from differences of sampling, since it falls within four times the P.E. of the original coefficient. Comparing the data further, we find that the syllable recognition test proved somewhat more difficult for the present group than for the group used in Part I. The mean and the S.D. of this test are both slightly lower in the present group, and there is a more pronounced piling up of scores at the lower end of the distribution. This condition might account in part for the difference in the reliability coefficients found in the two studies.

The reliability coefficient of the incidental memory test makes this test definitely unsatisfactory as a measure of individual differences. The distribution of scores is symmetrical enough as shown by the measures of skewness and kurtosis, but variability is extremely low. It is interesting to speculate on the possible causes of the unreliability of this test. In the literature on incidental memory (cf., for example (19), (27)), reliability coefficients are not reported. We cannot, therefore, compare the reliability coefficient of the present test with that of other incidental memory tests. Many of these tests, however, were shorter than our test, which, it will be remembered, contained 60 items. From the standpoint of accuracy of construction, preliminary experimentation, procedure, etc., our test compares favorably with tests already described. Hence, if reliability coefficients were available for

TABLE XII

EVALUATION OF THE TESTS*

Test	*Reliability Coefficient*	*Mean*	*Standard Deviation*	Sk/σ_{Sk}	*Ku.*
1. Analogies	.6983	60.3824	5.4065	+1.95	.2723
2. Memory for Movement	.8465	36.7294	6.4277	+3.43	.2603
3. Word-Word	.8543	20.0647	7.1753	−0.79	.2708
4. Syllable Recognition	.5256	45.8353	9.8037	−1.70	.3192
5. Logical Recognition	.8681	45.6765	10.0167	+1.03	.2809
6. Logical Recall	.8750	42.8471	10.7346	+0.05	.2942
7. Delayed Memory	.6377	7.7118	3.2834	−1.26	.2567
8. Incidental Memory	.3798	48.5353	3.2937	+0.76	.2757
9. Tonal Memory	.8768	38.2824	6.4343	+3.40	.2817
10. Pitch Discrimination	.8894	75.5706	10.3774	+5.12	.2418

*The means, standard deviations, and correlations of halves were computed by the Columbia University Statistical Bureau.

these other incidental memory tests, it is probable that they would be no higher than the reliability of our test. To be sure, an incidental memory test may have a low reliability coefficient simply because it *is* a measure of incidental memory. The fact that the subject's attention is directed away from the task of memorizing the items presented may increase the operation of chance factors upon the scores. Whatever the cause of the low reliability might be, the fact itself was deemed to be sufficient reason for our discarding as unsatisfactory the correlations obtained with this test.

The measurement of kurtosis shows that all of our distributions tend to be mesokurtic, as none of them deviates significantly from the kurtosis of the normal curve, i.e., .26315. With N = 170, the S.E. of our measure of kurtosis is .0213. Only three of the tests exhibit a significant degree of skewness, viz., memory for movement, in which the skewness is 3.43 times as large as its S.E., tonal memory in which it is 3.40, and pitch discrimination, in which it is 5.12. In all of these distributions, the piling up of scores at the upper end suggests that the tests are somewhat too easy for the present group. The probable effect of skewness upon our correlation coefficients is discussed further on p. 40 to 42.

II. *Intercorrelations of the Variables*

The correlations of the nine variables left after dropping the incidental memory test, with each other and with age, are reported in Table XIII. All except three of the correlations

with age are negative, but in all cases they are too low to have much effect upon the intercorrelations of the test variables. This indicates, of course, that our group was already very homogeneous in respect of age. The two highest correlations with age are those of logical recognition and of analogies, being —.1692 and —.1352, respectively. If we partial out age variability, the correlation between these two tests is reduced by .0130, viz., from .4322 to .4192. The effect of age is so slight, therefore, that as the intercorrelations were not to be employed further in the computation of tetrads, it was considered unnecessary to eliminate age variability. All coefficients have, however, been corrected for attenuation, and these corrected coefficients are given in Table XIV. The original numbering of the variables has been retained in this table, as well as in Table XIII, although number 8, incidental memory, is omitted.

One other factor already mentioned (p. 39) should now be considered for its possible effect upon the intercorrelations, namely, the significant skewness of three of the distributions. If the relation between our particular *tests* were all we wished to ascertain, then the correlation coefficients as computed would furnish an adequate measure of relationship. The correlation between two variables is correctly expressed by the product-moment correlation coefficient whether the distributions are skewed or normal (cf., for example (34), p. 481). But the correlations between the *traits* measured by our tests might have been higher if we had substituted more differentiating tests for those yielding skewed distributions. Hence,

Table XIII

INTERCORRELATIONS OF THE VARIABLES*

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*9*	*10*	*11*
1. Analogies		.0512	.0700	.1141	.4322	.3282	.0501	.0306	.1937	—.1352
2. Memory for Movement			.0295	.0611	.1100	.0675	—.0079	.2981	.2401	.0146
3. Word-Word				.2129	.1712	.1125	.0079	—.0298	.0118	.0172
4. Syllable Recognition					.1542	.0942	.0997	—.0053	.0678	—.0576
5. Logical Recognition						.6471	.0561	.0354	.2304	—.1692
6. Logical Recall							.0953	—.0091	.0771	—.1051
7. Delayed Memory								—.0280	—.0092	—.0428
9. Tonal Memory									.5115	.0540
10. Pitch Discrimination										—.0018
11. Age										

*These correlation coefficients were computed by the Columbia University Statistical Bureau.

TABLE XIV

INTERCORRELATIONS CORRECTED FOR ATTENUATION

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*9*	*10*
1. Analogies		.0666	.0906	.4883	.5551	.4199	.0751	.0391	.2458
2. Memory for Movement			.0347	.0916	.1283	.0781	—.0108	.3460	.2767
3. Word-Word				.3177	.4988	.1301	.0107	.0344	.0135
4. Syllable Recognition					.2283	.1389	.1722	—.0078	.0992
5. Logical Recognition						.7124	.0754	.0406	.0874
6. Logical Recall							.4276	—.0104	.0874
7. Delayed Memory								—.0374	—.0122
9. Tonal Memory									.5822
10. Pitch Discrimination									

in order to get a notion of the relationships between the traits themselves, rather than between the specific tests, we must consider the effect of skewness upon our correlations. It should be noted that none of our distributions exhibits any "jamming" of scores at either end. All of the distributions tail off gradually at both extremes, although the portions above and below the median are not symmetrical. Ordinarily, skewness lowers a correlation coefficient through curtailment of the range, i.e., through lowering the variability of the distribution. However, in our three distributions which have been curtailed through the use of too easy material, the S.D. is probably not reduced by more than 20% from what it would have been had the tests yielded normal distributions. This estimate, which is based upon data from our most highly skewed distribution (that of pitch discrimination in which sk = 5.12 $\sigma_{sk.}$), was arrived at by the following simple method. First, the actual range of scores above and below the median was expressed in terms of the P.E. of the skewed distribution. The scores in pitch discrimination, for example, extend from approximately —5 P.E. to exactly +2 P.E. Provided the half of the curve *below* the median has not been affected by the use of a too easy test, we may consider this half of the curve identical with that which would be obtained if the distribution were normal. This is a fair assumption if the addition of more difficult items at the upper end of the scale does not alter appreciably the scores made by individuals below the median. It follows, therefore, that 5 P.E. of our skewed distribution is approximately equal to 4 P.E. of the corresponding normal curve, since in a normal curve the range covered by all the

measures when N is 170 lies approximately between +4 P.E. and —4 P.E. (13, p. 104). The ratio between the estimated P.E. of the normal curve and the obtained P.E. of the skewed distribution is approximately 5/4. This, of course, is also the ratio of the S.D.'s of the two distributions.[1] Applying Pearson's formula[2] for estimating a correlation coefficient from the correlation obtained on a restricted range, we find that an obtained r of .10 would be raised to .12. The correction is even smaller in the case of lower r's. We can therefore conclude that in our correlation table, reductions in r resulting from skewness of the variables are not very large. We can certainly conclude that the large number of nearly zero correlations obtained cannot be attributed to skewness, since the reduction in the case of such low correlations must have been negligible.

The most outstanding feature about the correlations reported in Table XIV is the fact that the majority of them are so low. There are a few isolated high correlations, but the table as a whole does not indicate the presence of common factors. With 170 cases, a correlation coefficient must be approximately .20 in order to be over four times as large as its P.E. The P.E.[3] of the largest correlation in Table XIV, r_{56}, is .0310, that of the smallest positive correlation, r_{37}, is .0552. Hence it will be seen that very few of the correlations in the table can be regarded as significant deviations from zero. Five of the correlations are negative, four of these being correlations between *memory tests*. Furthermore, it is impossible to

---

[1] This rough method of estimating the ratio s_1/σ_1 was considered preferable in our case to the more elaborate procedure followed by Yerkes (33, p. 629-631) in which, after eliminating the "jammed" scores, the S.D. and r were computed on the remaining truncated distributions, and corresponding values were estimated therefrom for the total normal distributions. Since none of our distributions exhibits any jamming at either extreme, it would have been impossible to decide which class-intervals to cut off. If the entire upper half of the curve be cut off, the number of cases left is so small, and the error of sampling so large, that any results estimated from such computations are extremely dubious.

[2] Cf. Pearson (20), p. 23 and (22): $r_{12} = \dfrac{s_1}{\sigma_1} \dfrac{r_{12}}{\sqrt{1-\left(1-\dfrac{s_1^2}{\sigma_1^2}\right) r^2_{12}}}$

[3] Since these r's have been corrected for attenuation, their P.E.'s were computed by the following formula (13, p. 208-212):

$$P.E._{r_{\infty\infty}} = \frac{.6745 r_{\infty\infty}}{\sqrt{N}} \left\{ r^2_{\infty\infty} + \frac{1}{r^2_{12}} + \left(\frac{1}{4r^2_{11}} - \frac{r^2_{11}}{4} + r_{11} - 1\right) + \left(\frac{1}{4r^2_{2II}} - \frac{r^2_{2II}}{4} + r_{2II} - 1\right) \right\}^{\frac{1}{2}}$$

find *any set* of four tests in the table such that all the intercorrelations between them are significantly greater than zero, *even after correcting for attenuation.* For this reason, the tetrad criterion has not been applied to these data. The satisfaction of the tetrad criterion with data of this sort would tell us nothing regarding the organization of the traits involved and any conclusions drawn from such a procedure would be highly misleading. This point has been fully discussed elsewhere ((9) p. 257-260).

We shall now examine more fully some of the individual correlation coefficients, for any light that they may throw on the organization of our variables. The correlation between word-word and syllable recognition is .32 as compared with .33 in the 1930 study and .44 in Part I of the present monograph. This correlation, then seems to corroborate the earlier results obtained with these two tests. The delayed memory test yields the most consistently low correlations of any of the tests. Its correlations are remarkably low, even after correcting for attenuation.[4] This test, it will be remembered, is very similar in material and in method of administration to the earlier retained members test. The only differences between the two tests are that, in the delayed memory test, the length of the test has been cut in half, and recall is tested after a 48-hour interval instead of immediately. In view of the similarity between these two tests, a comparison of the correlations obtained with them may be of interest. In the 1930 study, retained members correlated .6066 with word-word; in Part I of the present study. the correlation was .6078; the delayed memory test, on the other hand, correlates only .0107 with word-word! This seems to suggest that if we alter a test even slightly, we may be changing its nature radically, so that it measures an entirely different ability. We shall return to this point in the conclusions at the end of this chapter.

The correlation of .35 between memory for movement and tonal memory is of considerable interest. These two tests are presented in different sense modalities and seem at first to be quite different, but upon further analysis of the procedure we find that very much the same techniques could be used by the subject in both tests. They both involve memory of a very

---

[4] It should be noted, however, that the correction for attenuation has less effect, the lower the r.

immediate sort, like the memory span, and retention is tested after each item in both tests; both involve remembering the *sequence* of very simple impressions; meaningful associations are very difficult to introduce in either test, and if used would probably be of relatively little assistance; finally both tests involve very accurate memory for *details*—the individual who retains the general idea of a situation and then fills in details would be at a disadvantage in both tests. It will be noticed that all of these common features in the two tests are specific devices of such a nature that if either test were altered, even slightly, they might no longer be of help. We may conclude that insofar as the tonal memory test is a test of memory at all, it measures the type of memory involved also in the memory for movement test. Its correlation with this test may be taken as an index of the contribution of memory to performance in a test of tonal memory. The correlation of .58 between tonal memory and pitch discrimination shows that the ability to deal with a certain type of material, the ability to discriminate between the various notes whose sequence is to be recalled, is even more important than memory in determining performance on the tonal memory test.[5] In order to show the relative contribution of memory and of pitch discrimination in the tonal memory test, the following regression equation was computed, in which x_9, tonal memory, is the dependent variable, and x_2 and x_{10}, memory for movement and pitch discrimination, are the independent variables.

$$x_9 = .1861x_2 + .4699x_{10}$$

In computing this regression equation, as well as the other regression equations reported further, raw correlation coefficients have been used, without correcting for attenuation. Kelley (13) p. 208, has pointed out that the use of corrected coefficients in a regression equation "would lead to a less close fit of the regression line and to a larger standard error of estimate of the criterion, knowing the independent variable, than occurs when the 'raw' correlation coefficient is used."[6]

[5] This correlation agrees quite closely with the correlations between the same two tests given in the literature. McCarthy (16) reports raw correlations of .56 and .41 in two different groups of students; Brown (5) reports a correlation of .52.

[6] The regression equation with corrected r's is reported herewith for comparative purposes:

$$x_9 = .2003x_2 + .5268x_{10}$$

The logical memory tests correlate very highly with each other, to the extent of .74. Turning to their correlations with other tests, we find that they correlate much higher with Analogies (.56 and .42 in the case of logical recognition and recall respectively) than with other memory tests. The correlation of logical recognition with word-word is only .20, and with syllable recognition .23; in the case of logical recall, the corresponding correlations are .13 and .14. Here we see the effect of test material to a most remarkable degree. We find two recognition tests, using different materials, correlating only .23, while a recognition and a recall test, using similar material, correlate .74. Facility in handling verbal material causes a correlation of .56, whereas the common feature of "retentiveness" in two tests produces a correlation as low as .13![7]

III. *Conclusions*

Our first conclusion is that the common factor previously found through a certain type of memory test, cannot be regarded as a general memory factor extending through *all* forms of memory. The large number of nearly zero and negative correlations obtained makes this conclusion inevitable. The low correlations obtained could not have arisen from any fault or limitation of technique, since, firstly, all the correlations were corrected for attenuation, secondly, the tests seem to discriminate very well between degrees of ability in the group of subjects used, as is shown by the measures of variability, and finally, all the distributions of memory test scores were normal, except in the case of tonal memory and memory for movement which showed a significant skewness—even in these two tests, the effect of the skewness on the correlation coefficients was shown to be negligible (cf. p. 42). Our results derive further validity from the fact that a few very high correlations *were* obtained, and that in two cases in which the correlations could be checked against the results of former studies they showed very close agreement.

A second conclusion is that here again, as in Part I, test material seems to have greater influence than test method in producing correlations. In addition, some evidence was pre-

[7] For a fuller analysis of the logical memory tests, the reader is referred to Part III, where certain additional control data will be discussed.

sented from several of our tests suggesting that the inter-test correlations might be due largely to the common applicability in more than one test situation of certain specific devices which play a large part in determining the subject's relative success or failure in a test.

PART III

Chapter VI

AN ANALYSIS OF LOGICAL RECALL AND RECOGNITION

I. *Construction of the Tests*

The main purpose of this section is to compare recall and recognition from various aspects. What is the *relation* between the two processes? Is recall intrinsically *more difficult* than recognition, when the two processes are measured by comparable tests? Let us examine briefly a few representative statements in the literature on recall and recognition, as well as the data upon which they are based. The distinction between the two processes was first emphasized by Wolfe (32). Following that, nearly all writers on the subject have referred to recall and recognition as more sharply differentiated—even contrasted—than other methods of testing memory such as, for instance, paired associates and retained members. The tendency has been to lump together a whole group of methods under the general term "recall" and to contrast these with *one* other method, "recognition." Hollingworth (11) in an early article refers to recall and recognition as involving the same neural patterns, but in opposite "directions." In recall, according to this explanation, the general setting is given and the "focal element" must be remembered; in recognition, the focal element is given and the general setting is to be remembered. Achilles (1) states in her conclusion (p. 73) that "The present study has been interested in the two methods of testing memory, recall and recognition. To reproduce or recall what one has seen or heard is different from recognizing it as something previously seen or heard when it is presented again. To the writer both experiences deserve the term memory, but the terms recall or reproduction and recognition should be used to distinguish them." What is the evidence upon which this distinction is based? If we refer to the inter-test correlations given by Achilles, we find that the average correlation between tests of recall and recognition for the same material is .21. The average correlation between all the recognition tests for different materials is .28 in one group and .37 in an-

other when a parallel set of tests is employed.[1] The correlations between recall tests for different materials are referred to by Achilles as averaging "about zero." Actually, these averages tell us very little, since the individual correlations vary widely, several being negative. What little they do show, however, does not support the view that recall and recognition tests are any more unrelated than, say, recall tests for different kinds of materials.

Lee (15) in a later study, also contrasts recall and recognition, and concludes in addition that differences in test content or material are less significant than differences in test "form." The only test "forms" compared are recall and recognition. The materials included words, nonsense, syllables, pictures of objects, and geometrical forms. Considering only the tests of immediate memory in Lee's study,[2] we find an average correlation of .18 between recall and recognition tests for similar material, .21 between recall tests for different materials, and .27 between recognition tests for different materials. The difference between any two average correlations is not large. Furthermore, in neither Lee's nor Achilles' study were reliability coefficients reported for any of the tests, so that we do not know to what extent the differences between correlations resulted from varying amounts of attenuation from errors of measurement.

Turning to a more recent study of memory, that of Bolton (4), in which reliability coefficients *were* considered, we find similar conclusions in the discussion and similar inconclusiveness in the data. The correlation between a word recognition test and a syllable recognition test reported by Bolton for 200 subjects was .457, the correlation between word recognition and a word recall test only .227. However, the reliability coefficients of these particular tests are .472 for word recall, .478 for word recognition, and .559 for syllable recognition. With errors of measurement looming so large in the results, it is difficult to draw valid conclusions.

There are, of course, investigators who dissent from the

[1] The difference of .07 points, between the r of .21 and that of .28 can hardly mean very much when a difference of .09 points (between .28 and .37) is produced by simply using parallel tests on a different sampling.

[2] Lee's delayed memory tests cannot be included in these averages since the identical materials were used as in the immediate memory tests, producing spuriously high correlations.

view that recall and recognition are distinct and independent processes. Stevenson Smith (28) opens a brief theoretical article on recognition and recall with the following words: "There is no definite difference between recall and recognition. This is the thesis that I shall try to defend" (p. 28). The article continues with suggestions on the qualitative similarities between the two processes. No data are reported save a few illustrations, but the viewpoint expressed is interesting in view of our quantitative findings on recall and recognition.

Our results suggest that the large differences often found between recall and recognition are owing largely to the specific testing procedure used. To take an extreme example, suppose we test recall by presenting a difficult prose passage and then requiring the subjects to write all they can remember of it, with no further nor more specific directions; and then suppose we test recognition by presenting the subjects with a number of statements on a similar passage read, with the instructions that they indicate which of these statements were contained in the passage read. In these two cases, we are measuring recall by one of the most difficult types of recall test available, and recognition, by the easiest kind of recognition test. Surely we cannot attribute the difference in scores forthwith to an intrinsic difference in the nature of the two processes.

In order to reduce as much as possible this error in the comparison of recall and recognition tasks, we have used more comparable methods of testing the two processes. Four tests were employed, two of which were the logical recall and recognition tests described in Part II, and the other two were parallel forms of these tests. Achilles (1) p. 67, suggests that the difference in central tendency between recall and recognition is *greater* for material relatively rich in associations. The evidence for this statement is that the difference in mean recall and recognition scores was greatest for proverbs, less for words, and still less for nonsense syllables. According to this principle, we should expect the greatest difference between recall and recognition to be found in such tests as our tests of logical memory, as they involve material rich in associations. This, in fact, was one of the chief factors determining our choice of material. We have some reason to expect, therefore, that any difference found between recall and recognition scores

with *our* tests, will be larger rather than smaller than that which would be obtained with other materials.

It will be remembered that the recall and recognition passages used in Part II were of the same length, and that each of the two test series consisted of 80 items. Recognition was tested by a five-alternative multiple choice test; 80 mimeographed statements, each followed by five alternatives, were presented to the subjects, who were to indicate the one correct alternative in each case. The fifth alternative in each item was "omitted," which meant either that the question had not been discussed in the passage read, or that none of the four alternatives given was the correct answer. The advantages of this method over the ordinary "plus-minus" method of testing recognition are, firstly, that it reduces greatly the chances of guessing correctly, and hence requires a smaller correction for guessing in the score. It is obviously preferable to reduce an error by the use of better controlled experimental technique, than to allow for it afterwards through large statistical "corrections." Secondly, the necessity of choosing from among five alternatives in each question makes the test a more difficult one, aside from the question of guessing, so that the procedure is somewhat more comparable to that in a recall test. The alternatives were selected so as to be equally plausible, and frequently the correct choice depended upon some very specific item in the passage read. Recall was tested by a series of very specific questions based upon the passage, the answer to each of which was to be a word or phrase. In the original forms of the test (form A), passage I, entitled "Camouflage in Nature," was given for recognition, and passage II, "The Architecture of the Ancient Dekka" for recall. Two parallel forms (form B) were constructed, in which passage I was used for recall, and passage II for recognition. The *presentation* of the passages was identical in both forms. The recall and recognition questions on each passage *dealt with the same details* in both tests, i.e., for almost every recognition question in form A, there was a corresponding recall question on the same passage in form B, and vice versa.

II. *Testing Procedure*

The results reported upon form A of both passages were secured in the study based upon 170 college women, reported in

Part II. Form B was given to a "control" group of 63 students in introductory psychology in Columbia University Extension Classes. It was not thought necessary to have a control group that was strictly comparable to the main group of 170, since comparisons were only made *within* each group. The mean recall and recognition scores were compared within each of the two groups. It was considered necessary to employ two groups, each of which took a different form of the tests, since the passages were not known to be equal in difficulty. To be sure, the two passages were constructed so as to be as similar as possible, but we did not consider it justifiable to *assume*, without any experimental verification, that equality had been achieved. If the recall scores were much lower than the recognition scores, it might indicate simply that the passage used in the recall test was more difficult in itself. If this were the case, the difference between the recall and recognition scores would be reduced or reversed in the other group of subjects taking form B, in which the recall and recognition passages were reversed. A further possible irregularity in the result arises from practice effect. The test which is given second may have an advantage because of the practice that the subjects have had in taking the first test. If helpful devices have been discovered in the first test, these would prove useful in the second, because of the similarity of the two tests. To allow for practice effect, the *order* of giving the recall and recognition tests was reversed in the two groups. A schedule of the testing in both groups is given in Table XV in order to clarify the control features used.

TABLE XV

TESTING SCHEDULE FOR RECALL AND RECOGNITION

Subjects	*Testing Period*	*Test*	*Passage Used*
Group I: 170 college students, female	I	Recognition, form A	"Camouflage in Nature"
	II	Recall, form A	"Architecture of the Ancient Dekka"
Group II: 63 college students, both sexes	I	Recall, form B	"Camouflage in Nature"
	II	Recognition, form B	"Architecture of the Ancient Dekka"

III. *Results and Conclusions*

The recall and recognition data on both groups are reported in Table XVI. Results are all expressed in terms of raw scores, i.e., total number of items correct with no correction for guessing. It is likely that many subjects *did* guess when not sure, and as a result, several items in the recognition test may be correct by chance only, a factor which could hardly operate in the recall test. In spite of this, the differences between recall and recognition scores are lower than might have been expected from the conclusions of previous studies on recall and recognition.

Table XVI
COMPARISON OF RECALL AND RECOGNITION

		Recall	*Recognition*
Group I: $N=170$	Mean	42.85	45.68
	S.D.	10.73	10.02
	σ_M	0.82	0.77
		$\sigma_{M1-M2}=.67$[3]	
Group II: $N=63$	Mean	34.67	44.59
	S.D.	12.52	12.88
	σ_M	1.58	1.62
		$\sigma_{M1-M2}=1.31$[3]	

Comparing the data secured on both groups, we find evidence that the two passages probably do not differ much in difficulty. The mean recognition scores in the two groups using different passages differ by only about one point. In the recall tests, the mean of Group I is approximately eight points higher than that of Group II. This latter difference might indicate a difference in the capacity of the subjects in the two groups, or a difference in the difficulty of the two passages. Both of these factors, however, operated also in the recognition test, in which the difference between the results on the two groups was practically negligible. A more plausible explanation of the difference is in terms of *practice effect,* since in Group II, the recall test was taken first, and was therefore at a disadvantage.

If we apply a correction for guessing to the mean scores in recognition, even the relatively slight differences found in

[3] The formula used in computing σ_{M1-M2} was:

$$\sigma_{M1-M2} = \sqrt{\sigma^2_{M1} + \sigma^2_{M2} - 2r_{12}\sigma_{M1}\,\sigma_{M2}}$$

favor of recognition disappear. The following formula, reported by Miller (18), has been used in correcting for guessing:

$$S = (S_1 - U) - \frac{nw}{n - 1}$$

in which,

S = corrected score
S_1 = number of items in the test
U = number of items omitted. This was taken as equal to 1.00, since that is the average number of omissions made by the subjects in each group
W = number of wrong items
N = number of alternatives used, which in our tests was 5.

Applying this correction to our recognition averages, we have:

	Corrected Recognition Average	*Recall Average (from Table XVI)*
Group I:	37.75	42.85
Group II:	35.99	34.67

The difference between recall and recognition is now practically negligible in Group II, and actually *reversed* in Group I, recall in that group having a slightly higher average than recognition. When all of the evidence is considered, it certainly seems that the alleged superiority of performance in recognition tests disappears as long as comparable measures are employed.

The fact that recall and recognition tests give approximately equal mean scores does not necessarily imply that they do not represent two relatively independent types of memory. The central tendencies may be equal, and still performance in the two tests may vary independently from one subject to another. We shall now examine this aspect of the comparison more fully. In Part II it was found that logical recall and recognition correlate with each other to the extent of .74. This in itself shows a very close relation between the two processes. The relationship can be further demonstrated by analyzing the factors which determine performance in the two tests, in other words, by comparing the correlations of recall and recognition with other tests. An examination of Table XIV (Part II, p. 41) reveals the fact that such correlations are very similar for recall and recognition. Regression equations showing the relative contribution of certain selected variables to per-

formance in logical recall and in logical recognition, furnish further evidence of the close similarity of the two processes. The following variables have been used in computing these regression equations:

x_5 = logical recognition	x_1 = analogies
x_6 = logical recall	x_3 = word-word
	x_4 = syllable recognition

The equation for logical recognition is:

$$x_5 = .4146x_1 + .1251x_3 + .0802x_4$$

The equation for logical recall is:

$$x_6 = .3179x_1 + .0816x_3 + .0406x_4$$

Uncorrected correlation coefficients have been used throughout in the computation of the regression equations.[4] It appears from these equations that the ability to handle verbal material, as measured by the analogies test, is much more important in determining achievement on the logical memory tests, than is the simple type of rote memory measured by the word-word and nonsense syllable recognition tests. The most striking fact about the two equations, however, is their similarity. The weights of the three independent variables in the two equations correspond closely. This offers further evidence of the close relationship between recall and recognition.

In conclusion, our results have shown first, that the alleged superiority of recognition over recall scores does not hold for logical memory, a type of memory in which one is led to expect the greatest difference between the two processes, according to previous writers. Secondly, the two processes are closely related, almost to the point of identity, when similar material is employed. This fact, together with the relatively large weight of analogies in the regression equations of both recall and recognition, again suggests the conclusion that material is more potent than method in determining inter-test correlations.

[4] For discussion of this procedure, cf. p. 44, Part II. The corresponding values in each equation, with r's corrected for attenuation, are given for comparative purposes:

$$x_5 = .5318x_1 + .1224x_3 + .0902x_4$$
$$x_6 = .4075x_1 + .0818x_3 + .0365x_4$$

Chapter VII

SUMMARY AND EVALUATION

In Part I of the present investigation, 140 college women were tested with four memory tests, two verbal tests, and two numerical tests. All of the tests had been used before, in the same form. The four memory tests dealt with a particular type of memory which may be described as immediate rote memory for visually presented material. In Part II, 170 subjects drawn from the same population as those in Part I, were given ten tests, most of which were new. Eight of the tests were memory tests, varying widely in all aspects save the common feature of retentivity. In Part III, an analysis of logical recall and recognition was undertaken, based upon some of the data from Study II, in conjunction with certain specially obtained control data. For the specific findings in each part, the reader is referred to the conclusions at the end of each section. In the present chapter, we shall bring together the various suggestions gleaned from each section, and endeavor to work out certain general implications of the results.

The chief conclusion that can be drawn from the present study is, clearly, that we cannot speak of a single common factor running through *all* forms of memory. The evidence from Part II is especially conclusive on this point. The very fact that in Part I, as well as in the 1930 study, the multiple correlation between our battery of memory tests and the factor common to those tests was so high, may be regarded as a precursor of our later findings. It seems to us that a factor which is measured so completely by a group of simple tests of a relatively narrow phase of memory, is not very likely to extend through a wide range of performance, not as likely at least, as if the multiple correlation had been lower, indicating that certain aspects of the common factor had not been touched upon by the tests used. The common factor previously found may have consisted almost entirely of certain special devices[1] which could be applied generally to rote memory for verbal

[1] We are using the terms "special device" and "technique" in a very similar sense to that in which Gates (10) uses these terms in his discussion of the nature of improvement due to training.

material. A concrete example may make this point clearer. When confronted with a series of disconnected words, many individuals, some more readily than others, will tend to group the words into a sentence or other more or less meaningful unit. This device is an undoubted advantage in any test involving memory of disconnected words. Differences in the speed, frequency, and skill with which different individuals group the words may produce variability in test scores through this factor alone. When changes in material, or in any other aspect of the tests, are introduced, as in Part II of our study, the same devices can no longer be of assistance in the tests, and the common factor disappears. Of course, all that our data can show directly is the *extent* of the common factor. Any statements regarding its *nature* are offered only as very tentative hypotheses. Reference has been made throughout this study to two probable interpretations of common factors: i.e., community of material and special techniques. These two interpretations are not mutually exclusive, but, on the contrary, the one *may* be explained in terms of the other. The evidence for either interpretation is admittedly meagre, as it was not our main purpose in this study to analyze the *nature* of the memory factor. It may be of some value nevertheless, to summarize at this point the facts that suggested each hypothesis.

First, let us consider the interpretation in terms of community of test *material.* The data of Achilles (1), Lee (15), and Bolton (4) on this question have been cited as inconclusive. The data in our 1930 study suggest that community of material is more significant in producing correlation than community of method, since the two pairs of tests which had to be pooled on account of exceptionally high correlations were both characterized by community of material. One of these test pools consisted of the word-word test and the retained members words tests; the other contained the picture-number and form-number tests. In Part I of the present study, likewise, the correlation between word-word and retained members was much higher than that between word-word and picture-number; in the former combination, material is similar and method differs, in the latter, method is identical but material differs. The very high correlation found in Part II between logical recall and recognition furnishes further evi-

dence of the effect of material. Using the two methods of testing memory which have usually been regarded as the most diverse, but applying them to materials which had been closely equated, we obtained with these two tests the highest correlation of any in the table. Finally, we may find some rather indirect evidence for our hypothesis in a comparison of Schneck's results on the verbal and numerical factors, both of which are *content* factors, with our results on the memory factor, a factor of *process*, or method. The tetrad equations in Schneck's study gave very clear-cut evidence of the presence of a single common factor through each of the two types of material, with no disturbing group factors through overlapping of method. In our 1930 study on the memory factor, on the other hand, we found many disturbing group factors from similarity of material, necessitating the pooling of tests and leaving a suggestion of minor group factors even in the final tetrad with the pooled tests. Throughout this discussion, we have purposely used the terms method and material, rather than form and content, because we wish to make the distinction in terms of the actual test situation, not in terms of the mental activity of the subject in dealing with that situation. The latter obviously cannot be discussed in a statistical study.

The second interpretation of group factors is based upon the common applicability to several test situations of certain special techniques which facilitate the subject's performance. If such devices account in large measure for inter-test correlations, we should expect large changes in such correlations when the procedure of the test is changed even slightly, as by introducing a delay. This is exactly what we do find, when we compare the correlations of the retained members test with the corresponding correlations of the delayed memory test, the correlations differing very markedly in the two cases. It will be remembered that these two tests were very similar except for the factor of delay. Certain devices which might facilitate the recall of a series of words immediately after their presentation might be of little or no value when recall is tested after 48 hours. For example, a preliminary survey of the subjects' responses in these two tests shows that in immediate recall, a large number of subjects tend to reproduce the words in the reverse order from the order of presentation, whereas in the delayed reproductions the original presentation order is

found more commonly. This fact suggests a possible difference in the nature of the two tests brought about by the delay alone. The effect of specific aids and devices is again suggested by the relatively high correlation found between tonal memory and memory for movement. Both tests involve a very immediate sort of memory for short series of discrete impressions, making possible the use of various common devices. Finally, all the evidence cited above as indicative of the greater potency of material than method, or process, in determining inter-test correlations, *might* be explained in terms of the use of common facilitating devices. It would seem that such devices could be applied more effectively and more widely within a given type of *material*, than within a given *process*. If the explanation of common factors in terms of such more or less widely applicable devices be correct, it suggests interesting possibilities regarding environmental influences and training in relation to problems of mental organization. Especially fruitful in this connection would be the experimental study of the effect of special training on inter-test correlations, as well as the comparison of correlation results in groups of widely diverse training and background.

BIBLIOGRAPHY

1. Achilles, E. M. Experimental Studies in Recall and Recognition. Arch. of Ps. No. 44, 1920. Pp. 80.
2. Anastasi, A. A Group Factor in Immediate Memory. Arch. of Ps. No. 120, 1930. Pp. 61.
3. Anderson, D. L. The Minnesota Mechanical Ability Tests. Pers. J. 6, 1928, 473-473.
4. Bolton, E. B. The Relation of Memory to Intelligence. J. Exp. Ps. 14, 1931, 37-67.
5. Brown, A. W. The Reliability and Validity of the Seashore Tests of Musical Talent. J. Appl. Ps. 12, 1928, 468-476.
6. Dunlap, J. W. and Kurtz, A. K. Handbook of Statistical Nomographs, Tables, and Formulas. World Book Co., 1932. Pp. 163.
7. Fisher, R. A. Statistical Methods for Research Workers. Lond. Oliver, 1930. Pp. 284.
8. Garrett, H. E. Statistics in Psychology and Education. N. Y. Long., 1929. Pp. 317.
9. Garrett, H. E. and Anastasi, A. The Tetrad Difference Criterion and the Measurement of Mental Traits. Ann. N. Y. Acad. Sci. 33, 1932, 233-282.
10. Gates, A. I. The Nature and Limit of Improvement Due to Training. 27th Year-Book of Natl. Soc. for the Study of Educ. Ch. 23, 441-460.
11. Hollingworth, H. L. Characteristic Differences between Recall and Recognition. Am. J. Ps. 24, 1913, 532-544.
12. Holzinger, K. J. On Scoring Multiple Response Tests. J. Ed. Ps. 15, 1924, 445-447.
13. Kelley, T. L. Statistical Method. N. Y. Mac., 1924. Pp. 390.
14. Kelley, T. L. Crossroads in the Mind of Man. Stanford Univ. Press, Calif., 1928. Pp. 238.
15. Lee, A. L. An Experimental Study of Retention and Its Relation to Intelligence. Ps. Rev. Mon. 34, 1925. Pp. 45.
16. McCarthy, D. A Study of the Seashore Measures of Musical Talent. J. Appl. Ps. 14, 1930, 437-455.
17. McGeoch, J. A. Memory. Ps. Bull., 1928, 25, 513-549.
18. Miller, G. F. Formulas for Scoring Tests in Which the Maximum Amount of Chance Is Known. Proc. Okla. Acad. Sci. 5, 1925, 30-42.
19. Myers, G. C. A Study in Incidental Memory. Arch. of Ps. 26, 1913. Pp. 108.
20. Pearson, Karl. Mathematical Contributions to the Theory of Evolution: XI. On the Influence of Natural Selection on the Variability and Correlation of Organs. Phil. Trans. of the Royal Soc. of London. Series A, vol. 200, 1903, 1-66.
21. Pearson, Karl. Mathematical Contributions to the Theory of Evolution: XIV. On the General Theory of Skew Correlation and Non-Linear Regression. Draper's Co. Res. Mem., Biometric Series, II, 1905. Pp. 54.
22. Pearson, Karl. On the General Theory of the Influence of Selection on Correlation and Variation. Biom. 8, 1911-12, 437-443.
23. Pintner, Rudolf. A Non-Language Group Intelligence Test. J. Appl. Ps. 3, 1919, 194-214.
24. Seashore, C. E. The Psychology of Musical Talent. Bost. Silver., 1919. Pp. 288.
25. Seashore, C. E. Manual of Instructions and Interpretations for Measures of Musical Talent. Columbia Graphophone Co., N. Y. C. Educ. Dept. Pp. 16.
26. Schneck, M. R. The Measurement of Verbal and Numerical Abilities. Arch. of Ps. 107, 1929. Pp. 49.

27. Shellow, S. M. Individual Differences in Incidental Memory. Arch. of Ps. 64, 1923. Pp. 53.
28. Smith, S. Recognition and Recall. Ps. Rev. 34, 1927, 28-33.
29. Spearman, C. The Abilities of Man. N. Y. Mac., 1927. Pp. 415.
30. Strong, E. K. The Effect of Length of Series upon Recognition Memory. Ps. Rev. 19, 1912, 447-462.
31. Strong, E. K. The Effect of Time Interval upon Recognition Memory. Ps. Rev. 20, 1913, 339-372.
32. Wolfe, H. K. Untersuchung über das Tongedächtniss. Phil. Stud. 3, 1886, 534-571.
33. Yerkes, R. M. Psychological Examining in the United States Army. Memoirs of the Nat. Acad. Sci. 15, 1921, 629-632.
34. Yule, G. U. On the Significance of Bravais' Formulae for Regression, etc., in the Case of Skew Correlation. Proc. of the Roy. Soc. of London, vol. 60, 1896-7, 477-489.

150.8
A673
no. 142
1932
Archives of psychology

ARCHIVES OF PSYCHOLOGY

List of numbers, continued from inside front cover

77. Tested Mentality as Related to Success in Skilled Trade Training: T. M. ABEL. $1.25.
78. Aggressive Behavior in a Small Social Group: E. M. RIDDLE. $1.75.
79. Memory Value of Advertisements: E. R. BRANDT. $1.25.
80. Critical Examination of Test-Scoring Methods: R. G. ANDERSON. $1.00.
81. Thermal Discrimination and Weber's Law: E. A. K. CULLER. $1.75.
82. Correlational Analysis of Typing Proficiency: L. ACKERSON. $1.50.
83. Recall as a Function of Perceived Relations: C. B. KEY. $1.25.
84. Consistency of Rate of Work: C. E. DOWD. $1.00.
85. Experimental Investigation of Recovery from Work: S. L. CRAWLEY. $1.25.
86. Facilitation and Inhibition: T. N. JENKINS. $1.00.
87. Variability of Performance in the Curve of Work: J. D. WEINLAND. $1.00.
88. Mental Hygiene Inventory: S. D. HOUSE. $1.50.
89. Mental Set and Shift: A. T. JERSILD. $1.25.
90. Experimental Investigation of Rest Pauses: C. W. MANZER. $1.25.
91. Routine and Varying Practice as Preparation for Adjustment to a New Situation: L. W. CRAFTS. $1.00.
93. Speed and Other Factors in "Racial" Differences: O. KLINEBERG. $1.50.
94. Relation of Reaction Time to Intelligence, Memory, and Learning: V. W. LEMMON. 80c.
95. Is the Latent Time in the Achilles Tendon Reflex a Criterion of Speed in Mental Reactions? G. H. ROUNDS. $1.25.
96. Predictive Value of Tests of Emotional Stability Applied to College Freshmen: E. G. FLEMMING. $1.00.
97. Vocabulary Information Test: A. L. WEEKS. $1.00.
98. Effect of Temporal Arrangements of Practice on the Mastery of an Animal Maze: S. A. COOK. 80c.
99. Recognition Time as a Measure of Confidence: G. H. SEWARD. $1.00.
100. Precision and Accuracy: G. W. HARTMAN. 80c.
101. Group Test of Home Environment: E. M. BURDICK. $1.50.
102. Effect of Material on Formal Syllogistic Reasoning: M. C. WILKINS. $1.25.
103. Effect of Incentives on Accuracy of Discrimination: H. C. HAMILTON. $1.25.
104. Validity of Norms with Special Reference to Urban and Rural Groups. M. E. SHIMBERG. $1.25.
105. Blood Pressure Changes in Deception: M. N. CHAPPELL. 80c.
106. Experimental Comparison of Psychophysical Methods: W. N. KELLOGG. $1.25.
107. Measurement of Verbal and Numerical Abilities: M. M. R. SCHNECK. $1.00.
108. Perseverative Tendency in Pre-School Children. A Study in Personality: H. M. CUSHING. $1.00.
109. Preliminary Study of the Effect of Training in Junior High School Shop Courses: L. D. ANDERSON. 80c.
110. Music Appreciation: An Experimental Approach to Its Measurement: M. J. ADLER. $1.50.
111. Motivation in Fashion: E. B. HURLOCK. $1.00.
112. Equality Judgments in Psychophysics: W. N. KELLOGG. $1.00.
113. Illusions in the Perception of Short Time Intervals: N. ISRAELI. 80c.
114. Further Studies of the Reading-Recitation Process in Learning: E. B. SKAGGS, S. GROSSMAN, L. KRUEGER, W. C. F. KRUEGER. 80c.
115. Factors Affecting the Galvanic Reflex: R. C. DAVIS. $1.00.
116. Infant's Feeding Reactions during the First Six Months: R. RIPIN. 80c.
117. Measurement of Mental Deterioration: H. BABCOCK. $1.25.
118. Phenomenon of Postural Persistence: L. S. SELLING. $1.00.
119. American Council on Education Rating Scale: Reliability, Validity, and Use: F. F. BRADSHAW. $1.00.
120. Group Factor in Immediate Memory: A. ANASTASI. $1.00.
121. Individual Differences in the Sense of Humor and Temperamental Differences: POLYXENIE KAMBOUROPOULOU. $1.00.
122. Suggestibility in Normal and Hypnotic States: G. W. WILLIAMS. $1.00.
123. Analytical Study of the Conditioned Knee-Jerk: G. R. WENDT. $1.25.
124. Race Differences in the Organization of Numerical and Verbal Abilities: J. W. DUNLAP. $1.25.
125. Errors of Measurement and Correlation: E. E. CURETON. $1.25.
126. Experience Factors Test Scores and Efficiency of Women Office Workers: N. BIRD. $1.00.
127. Delayed Reactions of Infants: A Sex Differences in Early Retentiveness: C. N. ALLEN. 80c.
128. Factors Measured by the Thorndike Intelligence Examination for High School Graduates: J. G. PEATMAN. $1.00.
129. Educational Success and Failure in Supernormal Children: J. REGENSBURG. $1.75.
130. Effect of Practice on the Visual Perception of Form: J. P. SEWARD, JR. $1.00.
131. Relation to College Grades of Some Factors other Than Intelligence: D. HARRIS. 80c.
132. Study of Psychological Differences Between "Racial" and National Groups in Europe: O. KLINEBERG. $1.00.
133. Emotional Differences of Delinquent and Non-Delinquent Girls of Normal Intelligence: A. COURTHIAL. $1.25.
134. Learning and Retention of Pleasant and Unpleasant Activities: H. CASON. $1.25.
135. Experimental Investigation of Brightness Constancy: R. B. MACLEOD. $1.25.
136. Rorschach Test as Applied to a Feeble-Minded Group: S. J. BECK. $1.00.
137. Retention after Intervals of Sleep and of Waking: E. B. van Ormer. $1.00.
138. Stimulus Temperature and Thermal Sensation: F. Heiser. $1.00.
139. Energy Cost Measurements on the Curve of Work: H. J. P. Schubert. $1.00.
140. Technique for the Measurement of Attitudes: R. Likert. $.80.
141. Speed Factor in Mental Tests: P. H. DuBois. ~~$1.00.~~ 50 c.
142. Further Studies on the Memory Factor: A. Anastasi. $1.00.

DIRECTORY OF
AMERICAN PSYCHOLOGICAL PERIODICALS

AMERICAN JOURNAL OF PSYCHOLOGY—Ithaca, N. Y.; Cornell University.
Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, Madison Bentley, and E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

JOURNAL OF GENETIC PSYCHOLOGY—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

PSYCHOLOGICAL REVIEW—Princeton, N. J.; Psychological Review Company.
Subscription $5.50. 540 pages annually. Edited by Howard C. Warren. Bi-monthly. General psychology. Founded 1894.

PSYCHOLOGICAL MONOGRAPHS—Princeton, N. J.; Psychological Review Company.
Subscription $6.00 per volume. 500 pages. Founded 1895. Edited by Herbert S. Langfeld. Published without fixed dates, each issue one or more researches.

PSYCHOLOGICAL INDEX—Princeton, N. J.; Psychological Review Company.
Subscription $4.00. 300-400 pages. Edited by Walter S. Hunter. An annual bibliography of psychological literature. Founded 1895.

PSYCHOLOGICAL BULLETIN—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 720 pages annually. Edited by Edward S. Robinson. Monthly (10 numbers). Psychological literature. Founded 1904.

ARCHIVES OF PSYCHOLOGY—New York, N. Y.; Columbia University.
Subscription $6.00. 500 pages per volume. Edited by R. S. Woodworth. Without fixed dates, each number a single experimental study. Founded 1906.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $5.00. 448 psges annually. Edited by Henry T. Moore. Quarterly. Abnormal and social. Founded 1906.

PSYCHOLOGICAL CLINIC—Philadelphia, Pa.; Psychological Clinic Press.
Subscription $3.00. 288 pages. Edited by Lightner Witner. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

PSYCHOANALYTIC REVIEW—Washington, D. C.: 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Edited by W. A. White and S. E. Jelliffe. Quarterly. Psychoanalysis. Founded 1913.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY—Princeton, N. J.
Psychological Review Company, 700 pages annually. Experimental. Subscription $7.00. Founded 1916. Bi-monthly. Edited by S. W. Fernberger.

JOURNAL OF APPLIED PSYCHOLOGY—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.50. 400 pages annually. Edited by James P. Porter. Bi-monthly. Founded 1917.

JOURNAL OF COMPARATIVE PSYCHOLOGY—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

COMPARATIVE PSYCHOLOGY MONOGRAPHS—Baltimore, Md.; The Johns Hopkins Press.
Subscription $5.00. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.

GENETIC PSYCHOLOGY MONOGRAPHS—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

PSYCHOLOGICAL ABSTRACTS—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $6.00. 700 pages annually. Edited by Walter S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

JOURNAL OF GENERAL PSYCHOLOGY—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually (2 volumes.) Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

JOURNAL OF SOCIAL PSYCHOLOGY—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500 pages annually. Edited by John Dewey, Carl Murchison, and international cooperating board. Quarterly. Political, racial, and differential psychology. Founded 1929.